41/pts

MICROBIAL β-GLUCURONIDASE GENES, GENE PRODUCTS AND USES THEREOF

TECHNICAL FIELD

The present invention relates generally to microbial β-glucuronidases, and more specifically to secreted forms of β-glucuronidase, and uses of these β-glucuronidases.

BACKGROUND OF THE INVENTION

The enzyme β-glucuronidase (GUS; E.C.3.2.1.31) hydrolyzes a wide variety of glucuronides. Virtually any aglycone conjugated to D-glucuronic acid through a β-O-glycosidic linkage is a substrate for GUS. In vertebrates, glucuronides containing endogenous as well as xenobiotic compounds are generated through a major detoxification pathway and excreted in urine and bile.

Escherichia coli, the major organism resident in the large intestine of vertebrates, utilizes the glucuronides generated in the liver and other organs as an efficient carbon source. Glucuronide substrates are taken up by *E. coli* via a specific transporter, the glucuronide permease (U.S. Patent No. 5,288,463 and 5,432,081), and cleaved by β-glucuronidase, releasing glucuronic acid residues that are used as a carbon source. In general, the aglycone component of the glucuronide substrate is not used by *E. coli* and passes back across the bacterial membrane into the gut to be reabsorbed into the bloodstream and undergo glucuronidation in the liver, beginning the cycle again. In *E. coli*, β-glucuronidase is encoded by the gusA gene (Novel and Novel, *Mol. Gen. Genet. 120*:319-335, 1973), which is one member of an operon comprising two other protein-encoding genes, gusB encoding a permease (PER) specific for β-glucuronides, and gusC encoding an outer membrane protein (OMP) that facilitates access of glucuronides to the permease located in the inner membrane.

While β-glucuronidase activity is expressed in almost all tissues of vertebrates and their resident intestinal flora, GUS activity is absent in most other

organisms. Notably, plants, most bacteria, fungi, and insects are reported to largely, if not completely, lack GUS activity. Thus, GUS is ideal as a reporter molecule in these organisms and has become one of the most widely used reporter systems for these organisms.

In addition, because both endogenous and xenobiotic compounds are generally excreted from vertebrates as glucuronides, β-glucuronidase is widely used in medical diagnostics, such as drug testing. In therapeutics, GUS has been used as an integral component of prodrug therapy. For example, a conjugate of GUS and a targeting molecules, such as an antibody specific for a tumor cell type, is delivered along with a nontoxic prodrug, provided as a glucuronide. The antibody targets the cell and GUS cleaves the prodrug, releasing an active drug at the target site.

Because the *E. coli* GUS enzyme is much more active and stable than the mammalian enzyme against most biosynthetically derived ß-glucuronides (Tomasic and Keglevic, *Biochem J 133*:789, 1973; Levvy and Conchie, 1966), the *E. coli* GUS is preferred in both reporter and medical diagnostic systems.

Production of GUS for use in *in vitro* assays, such as medical diagnostics, however, is costly and requires extensive manipulation as GUS must be recovered from cell lysates. A secreted form of GUS would reduce manufacturing expenses, however, attempts to cause secretion have been largely unsuccessful. In addition, for use in transgenic organisms, the current GUS system has somewhat limited utility because enzymatic activity is detected intracellularly by deposition of toxic colorimetric products during the staining or detection of GUS. Moreover, in cells that do not express a glucuronide permease, the cells must be permeabilized or sectioned to allow introduction of the substrate. Thus, this conventional staining procedure generally results in the destruction of the stained cells. In light of these limitations, a secreted GUS would facilitate development of non-destructive marker systems, especially useful for agricultural field work.

Furthermore, the *E. coli* enzyme, although more robust than vertebrate GUS, has characteristics that limit its usefulness. For example, it is heat-labile and

inhibited by detergents and end product (glucuronic acid). For many applications, a more resilient enzyme would be beneficent.

The present invention provides gene and protein sequences of microbial β-glucuronidases, variants thereof, and use of the proteins as a transformation marker, effector molecule, and component of medical diagnostic and therapeutic systems, while providing other related advantages.

SUMMARY OF INVENTION

In one aspect, an isolated nucleic acid molecule is provided comprising a nucleic acid sequence encoding a microbial of β-glucuronidase, provided that the β-glucuronidase is not from *E. coli*. Nucleic acid sequences are provided for β-glucuronidases from Thermotoga, *Staphylococcus*, Staphylococcus, Salmonella, Enterobacter, and Pseudomonas. In certain embodiments, the nucleic acid molecule encoding β-glucuronidase is derived from a eubacteria, such as purple bacteria, gram(+) bacteria, cyanobacteria, spirochaetes, green sulphur bacteria, bacteroides and flavobacteria, planctomyces, chlamydiae, radioresistant micrococci, and thermotogales.

In another aspect, microbial β-glucuronidases are provided that have enhanced characteristics. In one aspect, thermostable β-glucuronidases and nucleic acids encoding them are provided. In general, a thermostable β-glucuronidase has a half-life of at least 10 min at 65°C. In preferred embodiments, the thermostable β-glucuronidase is from Thermotoga or *Staphylococcus* groups. In other embodiments, the β-glucuronidase converts at least 50 nmoles of p-nitrophenyl-glucuronide to p-nitrophenyl per minute, per microgram of protein. In even further embodiments, the β-glucuronidase retains at least 80% of its activity in 10 mM glucuronic acid.

In another aspect, fusion proteins of microbial β-glucuronidase or an enzymatically active portion thereof are provided. In certain embodiments, the fusion partner is an antibody or fragment thereof that binds antigen.

In other aspects, expression vectors comprising a gene encoding a microbial β-glucuronidase or a portion thereof that has enzymatic activity in operative linkage with a heterologous promoter are provided. In such a vector, the microbial β-

glucuronidase is not *E. coli* β-glucuronidase. In the expression vectors, the heterologous promoter is a promoter selected from the group consisting of a developmental type-specific promoter, a tissue type-specific promoter, a cell type-specific promoter and an inducible promoter. The promoter should be functional in the host cell for the expression vector. Examples of cell types include a plant cell, a bacterial cell, an animal cell and a fungal cell. In certain embodiments, the expression vector also comprises a nucleic acid sequence encoding a product of a gene of interest or portion thereof. The gene of interest may be under control of the same or a different promoter.

Isolated forms of recombinant microbial β-glucuronidase are also provided in this invention, provided that the microbial β-glucuronidase is not *E. coli* β-glucuronidase. The recombinant β-glucuronidases may be from eubacteria, archaea, or eucarya. When eubacteria β-glucuronidases are clones, the eubacteria is selected from purple bacteria, gram(+) bacteria, cyanobacteria, spirochaetes, green sulphur bacteria, bacteroides and flavobacteria, planctomyces, chlamydiae, radioresistant micrococci, and thermotogales and the like.

The present invention also provides methods for monitoring expression of a gene of interest or a portion thereof in a host cell, comprising: (a) introducing into the host cell a vector construct, the vector construct comprising a nucleic acid molecule according to claim 1 and a nucleic acid molecule encoding a product of the gene of interest or a portion thereof; (b) detecting the presence of the microbial β-glucuronidase, thereby monitoring expression of the gene of interest; methods for transforming a host cell with a gene of interest or portion thereof, comprising: (a) introducing into the host cell a vector construct, the vector construct comprising a nucleic acid sequence encoding a microbial β-glucuronidase, provided that the microbial β-glucuronidase is not *E. coli* β-glucuronidase, and a nucleic acid sequence encoding a product of the gene of interest or a portion thereof, such that the vector construct integrates into the genome of the host cell; and (b) detecting the presence of the microbial β-glucuronidase, thereby establishing that the host cell is transformed.

Methods are also provided for positive selection for a transformed cell, comprising: (a) introducing into a host cell a vector construct, the vector construct comprising nucleic acid sequence encoding a microbial β-glucuronidase, provided that the microbial β-glucuronidase is not *E. coli* β-glucuronidase; (b) exposing the host cell to the sample comprising a glucuronide, wherein the glucuronide is cleaved by the β-glucuronidase, such that the compound is released, wherein the compound is required for cell growth. In all these methods, a microbial glucuronide permease gene may be also introduced.

Transgenic plants expressing a microbial β-glucuronidase other than *E. coli* β-glucuronidase are also provided. The present invention also provides seeds of transgenic plants. Transgenic animals, such as aquatic animals are also provided. Methods for identifying a microorganism that secretes β-glucuronidase, are provided comprising: (a) culturing the microorganism in a medium containing a substrate for β-glucuronidase, wherein the cleaved substrate is detectable, and wherein the microorganism is an isolate of a naturally occurring microorganism or a transgenic microorganism; and (b) detecting the cleaved substrate in the medium. In certain embodiments, the microorganism is cultured under specific conditions that are favorable to particular microorganisms.

In another aspect, a method for providing an effector compound to a cell in a transgenic plant is provided. The method comprises (a) growing a transgenic plant that comprises an expression vector, comprising a nucleic acid sequence encoding a microbial β-glucuronidase in operative linkage with a heterologous promoter and a nucleic acid sequence comprising a gene encoding a cell surface receptor for an effector compound and (b) exposing the transgenic plant to a glucuronide, wherein the glucuronide is cleaved by the β-glucuronidase, such that the effector compound is released. This method is especially useful for directing glucuronides to particular and specific cells by further introducing into the transgenic plant a vector construct comprising a nucleic acid sequence that binds the effector compound. The effector compound can then be used to control expression of a gene of interest by linking a gene of interest with the nucleic acid sequence that binds the effector compound.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition, various references are set forth below which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.), and are therefore incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents DNA sequence of an approximately 6 kb fragment that encodes β-glucuronidase from *Staphylococcus*.

Figure 2 is a schematic of the DNA sequence of a *Staphylococcus* 6 kb fragment showing the location and orientation of the major open reading frames. S-GUS is β-glucuronidase.

Figures 3A-B present amino acid sequences of representative microbial β-glucuronidases.

Figures 4A-J present DNA sequences of representative microbial β-glucuronidases.

Figures 5A-C present amino acid alignments of *Staphylococcus* GUS (SGUS) *E. coli* GUS (EGUS) and human GUS (HGUS)(5A). Microbial GUSes (5B) and nucleotide sequence alignments of *Staphylococcus, Salmonella,* and *Pseudomonas* β-glucuronidases.

Figure 6 is a graph showing that *Staphylococcus* GUS is secreted in *E. coli* transformed with an expression vector encoding *Staphylococcus* GUS. The secretion index is the percent of total activity in periplasm less the percent of total β-galactosidase activity in periplasm.

Figure 7 is a graph illustrating the half-life of *Staphylococcus* GUS and *E. coli* GUS at 65°C.

Figure 8 is a graph showing the turnover number of *Staphylococcus* GUS and *E. coli* GUS enzymes at 37°C.

Figure 9 is a graph showing the turnover number of *Staphylococcus* GUS and *E. coli* GUS enzymes at room temperature.

Figure 10 is a graph presenting relative enzyme activity of *Staphylococcus* GUS in various detergents.

Figure 11 is a graph presenting relative enzyme activity of *Staphylococcus* GUS in the presence of glucuronic acid.

Figure 12 is a graph presenting relative enzyme activity of *Staphylococcus* GUS in various organic solvents and in salt.

Figures 13A-C present a DNA sequence of *Staphylococcus* GUS that is codon-optimized for production in *E. coli*.

Figure 14 is a schematic of the DNA sequence of *Staphylococcus* GUS that is codon-optimized for production in *E. coli*.

Figure 15 presents schematics of two expression vectors for use in yeast (upper figure) and plants (lower figure).

Figure 16 is a DNA sequence of a *Salmonella* gene β--glucuronidase.

Figure 17 is an amino acid sequence of a *Salmonella* gene β--glucuronidase translated from the DNA sequence.

Figure 18A-C presents an alignment of amino acids of three β--glucuronidase gene products: Staph (*Staphylococcus*), E. coli, Sal (a *Salmonella*).

Figure 19A-G presents an alignment of nucleotides of three β--glucuronidases; Staph (*Staphylococcus*), E. coli, Sal (*Salmonella*).

DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to set forth definitions of certain terms that will be used hereinafter.

As used herein, "β-glucuronidase" refers to an enzyme that catalyzes the hydrolysis of β-glucuronides. Assays and some exemplary substrates for determining β--glucuronidase activity, also known as GUS activity, are provided in U.S. Patent No. 5,268,463. In assays to detect β-glucuronidase activity, fluorogenic or chromogenic substrates are preferred. Such substrates include, but are not limited to, p-nitrophenyl β-D-glucuronide and 4-methylumbelliferyl β-D-glucuronide.

As used herein, a "secreted form of a microbial β-glucuronidase" refers to a microbial β-glucuronidase that is capable of being localized to an extracellular environment of a cell, including extracellular fluids, periplasm, or is membrane bound on the external face of a cell but is not an integral membrane protein. Some of the protein may be found intracellularly. The amino acid and nucleotide sequences of exemplary secreted β-glucuronidases are presented in Figures 1 and 16 and SEQ ID Nos.: 1, 2, _____ and _____. Secreted microbial GUS also encompasses variants of β-glucuronidase. A variant may be a portion of the secreted β-glucuronidase and/or have amino acid substitutions, insertions, and deletions, either found naturally as a polymorphic allele or constructed. A variant may also be a fusion of all or part of GUS with another protein.

As used herein, "percent sequence identity" is a percentage determined by the number of exact matches of amino acids or nucleotides to a reference sequence divided by the number of residues in the region of overlap. Within the context of this invention, preferred amino acid sequence identity for a variant is at least 75% and preferably greater than 80%, 85%, 90% or 95%. Such amino acid sequence identity may be determined by standard methodologies, including use of the National Center for Biotechnology Information BLAST search methodology available at www.ncbi.nlm.nih.gov. The identity methodologies preferred are non-gapped BLAST. However, those described in U.S. Patent 5,691,179 and Altschul *et al., Nucleic Acids Res.* 25:3389-3402, 1997, all of which are incorporated herein by reference, are also useful. Accordingly, if Gapped BLAST 2.0 is utilized, then it is utilized with default settings. Further, a nucleotide variant will typically be sufficiently similar in sequence to hybridize to the reference sequence under stringent hybridization conditions (for nucleic acid molecules over about 500 bp, stringent conditions include a solution comprising about 1 M Na+ at 25° to 30°C below the Tm; *e.g.*, 5 x SSPE, 0.5% SDS, at 65°C; *see*, Ausubel, *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing, 1995; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989). Some variants may not hybridize to the reference sequence because of codon degeneracy, such as degeneracies introduced for codon optimization in a

particular host, in which case amino acid identity may be used to assess similarity of the variant to the reference protein.

As used herein, a "glucuronide" or "β-glucuronide" refers to an aglycone conjugated in a hemiacetal linkage, typically through the hydroxyl group, to the C1 of a free D-glucuronic acid in the β configuration. Glucuronides include, but are not limited to, O-glucuronides linked through an oxygen atom, S-glucuronides, linked through a sulfur atom, N-glucuronides, linked through a nitrogen atom and C-glucuronides, linked through a carbon atom (*see,* Dutton, *Glucuronidation of Drugs and Other Compounds,* CRC Press, Inc. Boca Raton, FL pp13-15). β-glucuronides consist of virtually any compound linked to the C1-position of glucuronic acid as a beta anomer, and are typically, though by no means exclusively, found as an O-glycoside. β-glucuronides are produced naturally in most vertebrates through the action of UDP-glucuronyl transferase as a part of the process of solubilizing, detoxifying, and mobilizing both natural and xenobiotic compounds, thus directing them to sites of excretion or activity through the circulatory system.

β-glucuronides in polysaccharide form are also common in nature, most abundantly in vertebrates, where they are major constituents of connective and lubricating tissues in polymeric form with other sugars such as N-acetylglucosamine (*e.g.*, chondroitan sulfate of cartilage, and hyaluronic acid, which is the principle constituent of synovial fluid and mucus). Other polysaccharide sources of β -glucuronides occur in bacterial cell walls, *e.g.*, cellobiuronic acid. β-glucuronides are relatively uncommon or absent in plants. Glucuronides and galacturonides found in plant cell wall components (such as pectin) are generally in the alpha configuration, and are frequently substituted as the 4-O-methyl ether; hence, such glucuronides are not substrates for β-glucuronidase.

An "isolated nucleic acid molecule" refers to a polynucleotide molecule in the form of a separate fragment or as a component of a larger nucleic acid construct, that has been separated from its source cell (including the chromosome it normally resides in) at least once in a substantially pure form. Nucleic acid molecules may be

comprised of a wide variety of nucleotides, including DNA, RNA, nucleotide analogues, have protein backbones (*e.g.*, PNA) or some combination of these.

Microbial β-glucuronidase genes

As noted above, this invention provides gene sequences and gene products for microbial β-glucuronidases including secreted β-glucuronidases. As exemplified herein, genes from microorganisms, including genes from *Staphylococcus* and *Salmonella* that encode a secreted β-glucuronidase, are identified and characterized biochemically, genetically, and by DNA sequence analysis. Exemplary isolations of β-glucuronidase genes and gene products from several phylogenetic groups, including *Staphylococcus*, *Thermotoga*, *Pseudomonas*, *Salmonella*, *Staphylococcus*, *Enterobacter*, *Arthobacter*, and the like, are provided herein. Microbial β--glucuronidases from additional organisms may be identified as described herein or by hybridization of one of the microbial β-glucuronidase gene sequence to genomic or cDNA libraries, by genetic complementation, by function, by amplification, by antibody screening of an expression library and the like (*see* Sambrook *et al.*, *infra* Ausubel *et al.*, *infra* for methods and conditions appropriate for isolation of a β-glucuronidase from other species).

The presence of a microbial β-glucuronidase may be observed by a variety of methods and procedures. Particularly useful screens for identifying β--glucuronidase are biochemical screening and genetic complementation. Test samples containing microbes, may be obtained from sources such as soil, animal or human skin, saliva, mucous, feces, water, and the like. Microbes present in such samples include organisms from the phylogenetic domains, *Eubacteria*, *Archaea*, and *Eucarya* (Woese, *Microbiol. Rev. 58*: 1-9, 1994), the Eubacteria phyla: purple bacteria (including the α, β, γ, and δ subdivisions), gram (+) bacteria (including the high G+C content, low G+C content, and photosynthetic subdivisions), cyanobacteria, spirochaetes, green sulphur bacteria, bacteroides and flavobacteria, planctomyces and relatives, chlamydiae, radioresistant micrococci and relatives, and thermotogales. It will be appreciated by those in the art that the names and number of the phyla may vary somewhat according

to the precise criteria for categorization (*see* Strunk *et al.*, *Electrophoresis 19*: 554, 1998). Other microbes include, but are not limited to, entamoebae, fungi, and protozoa.

Colonies of microorganisms are generally obtained by plating on a suitable substrate in appropriate conditions. Conditions and substrates will vary according to the growth requirements of the microorganism. For example, anaerobic conditions, liquid culture, or special defined media may be used to grow the microorganisms. Many different selective media have been devised to grow specific microorganisms (*see, e.g*, Merck Media Handbook). Substrates such as deoxycholate, citrate, etc. may be used to inhibit extraneous and undesired organisms such as gram-positive cocci and spore forming bacilli. Other substances to identify particular microbes (*e.g.*, lactose fermenters, gram positives) may also be used. A glucuronide substrate is added that is readily detectable when cleaved by β-glucuronidase. If GUS is present, the microbes will stain; a microbe that secretes β-glucuronidase should exhibit a diffuse staining (halo) pattern surrounding the colony.

A complementation assay may be additionally performed to verify that the staining pattern is due to expression of a GUS gene or to assist in isolating and cloning the GUS gene. Briefly, in this assay, the candidate GUS gene is transfected into an *E. coli* strain that is deleted for the GUS operon (*e.g.*, KW1 described herein), and the staining pattern of the transfectant is compared to a mock-transfected host. For isolation of the GUS gene by complementation, microbial genomic DNA is digested by *e.g.*, restriction enzyme reaction and ligated to a vector, which ideally is an expression vector. The recombinants are then transfected into a host strain, which ideally is deleted for endogenous GUS gene (*e.g.*, KW1). In some cases, the host strain may express GUS gene but preferably not in the compartment to be assayed. If GUS is secreted, the transfectant should exhibit a diffuse staining pattern (halo) surrounding the colony, whereas, the host will not.

The microorganisms can be identified in myriad ways, including morphology, virus sensitivity, sequence similarity, metabolism signatures, and the like. A preferred method is similarity of rRNA sequence determined after amplification of genomic DNA. A region of rRNA is chosen that is flanked by conserved sequences that

will anneal a set of amplification primers. The amplification product is subjected to DNA sequence analysis and compared to known rRNA sequences described.

In one exemplary screen, a bacterial colony isolated from a soil sample displays a strong, diffuse staining pattern. The bacterium was originally identified as a *Staphylococcus* by sequence determination of 16S rRNA after amplification. Additional 16S sequence information shows that this bacterium is a *Staphylococcus*. A genomic library from this bacterium is constructed in the vector pBSII KS+. The recombinant plasmids are transfected into KW1, a strain deleted for the β-glucuronidase operon. One resulting colony, containing the plasmid pRAJa17.1, exhibited a strong, diffuse staining pattern similar to the original isolate.

In other exemplary screens of microorganisms found in soil and in skin samples, numerous microbes exhibit a diffuse staining pattern around the colony or stained blue. The phylogenetic classifications of some of these are determined by sequence analysis of 16S rRNA. At least eight different genera are represented. Genetic complementation assays demonstrate that the staining pattern is most likely due to expression of the GUS gene. Not all complementation assays yield positive results, however, which may be due to the background genotype of the receptor strain or to restriction enzyme digestion within the GUS gene. The DNA sequence and predicted amino acid sequences of the GUS genes from several of these microorganisms found in these screens microorganisms are determined.

A DNA sequence of the GUS gene contained in the insert of pRAJa17.1 is presented in Figure 1 and as SEQ ID No: ___. A schematic of the insert is presented in Figure 2. The β-glucuronidase gene contained in the insert is identified by similarity of the predicted amino acid sequence of an open reading frame to the *E. coli* and human β-glucuronidase amino acid sequences (Figure 5A). Overall, *Staphylococcus* β--glucuronidase has approximately 47-49% amino acid identity to *E. coli* GUS and to human GUS. An open reading frame of *Staphylococcus* GUS is 1854 bases, which would result in a protein that is 618 amino acids in length. The first methionine codon, however, is unlikely to encode the initiator methionine. Rather the second methionine codon is most likely the initiator methionine. Such a translated product is 602 amino

acids long and is the sequence presented in Figures 3A-B and 4A-I. The assignment of the initiator methionine is based upon a consensus Shine-Dalgarno sequence found upstream of the second Met, but not the first Met, and alignment of the *Staphylococcus*, human, and *E. coli* GUS amino acid sequences. Furthermore, as shown herein, *Staphylococcus* GUS gene lacking sequence encoding the 16 amino acids is expressed in *E. coli* transfectants. In addition, the 16 amino acids (Met-Leu-Ile-Ile-Thr-Cys-Asn-His-Leu-His-Leu-Lys-Arg-Ser-Ala-Ile) SEQ ID No. ____ are not a canonical signal peptide sequence.

There is a single Asn-Asn-Ser sequence (residues 118-120 in Figures 3A-B) that can serve as a site for N-glycosylation in the ER. Furthermore, unlike the *E. coli* and human β-glucuronidases, which have 9 and 4 cysteines respectively, the *Staphylococcus* protein has only a single Cys residue (residue 499 in Figures 3A-B).

Two GUS sequences from *Salmonella* are analysed and found to be identical. The nucleotide sequence and its amino acid translate are shown in Figs 16 and 17. There are 7 cysteines and a single glycosylation site (Asn-Leu-Ser) at residue 358 (referenced to the *E. coli* sequence). Amino acid alignments are shown in Figure 18 and nucleotide alignments in Figure 19. *Salmonella* GUS has 71% nucleotide identity to *E. coli*, 51% to *Staphylococcus* and 85% amino acid identity to *E. coli* and 46% to *Staphylococcus*.

The DNA sequences of GUS genes from *Staphylococcus homini*, *Staphylococcus warneri*, *Thermotoga maritima* (TIGR *Thermotoga* database), *Enterobacter*, *Salmonella*, and *Pseudomonas* are presented in Figures 4A-J and SEQ ID Nos. ____. Predicted amino acid sequences are shown in Figures 3A-B and SEQ ID Nos. _____. The amino acid sequences are shown in alignment in Figures 5A-C. The signature peptide sequences for glycosyl hydrolases (Henrissat, *Biochem Soc Trans 26*:153, 1998; Henrissat B *et al.*, *FEBS Lett 27*:425, 1998) are located from amino acids 333 to 358 and from amino acids 406 to 420 (*Staphylococcus* numbering in Figures 3A and 5B). The catalytic nucleophile is Glu 344 (*Staphylococcus* numbering) (Wong *et al.*, *J. Biol Chem. 18*: 34057, 1998). Within these two signature regions, 17/26 and 8/15 residues are identical across the six presented sequences. At the non-identical positions,

most of the sequences share an identical residue. Thus, the sequences are highly conserved in these regions (identity between *Staphylococcus* and each other GUS gene ranges from 65% to 100% in signature 1 and from 73% to 100% in signature 2) (*see* Figure 5B). In contrast, between *Staphylococcus* and β-galactosidase, another glycosyl hydrolase that has signature sequences, identity is 46% in signature 1 and 73% in signature 2.

In addition, portions or fragments of microbial GUS may be isolated or constructed for use in the present invention. For example, restriction fragments can be isolated by well-known techniques from template DNA, *e.g.*, plasmid DNA, and DNA fragments, including, but limited to, digestion with restriction enzymes or amplification. Furthermore, oligonucleotides of 12 to 100 nt, 12 to 50 nt, 15 to 50 nt, can be synthesized or isolated from recombinant DNA molecules. One skilled in the art will appreciated that other methods are available to obtain DNA or RNA molecules having at least a portion of a microbial GUS sequence. Moreover, for particular applications, these nucleic acids may be labeled by techniques known in the art, such as with a radiolabel (*e.g.*, ^{32}P, ^{33}P, ^{35}S, ^{125}I, ^{131}I, ^{3}H, ^{14}C), fluorescent label (*e.g.*, FITC, Cy5, RITC, Texas Red), chemiluminescent label, enzyme, biotin and the like.

In certain aspects, the present invention provides fragments of microbial GUS genes. Fragments may be at least 12 nucleotides long (*e.g.*, at least 15 nt, 17 nt, 20 nt, 25 nt, 30 nt, 40 nt, 50 nt). Fragments may be used in hybridization methods (*see*, exemplary conditions described infra) or inserted into an appropriate vector for expression or production. In certain aspects, the fragments have sequences of one or both of the signatures or have sequence from at least some of the more highly conserved regions of GUS (*e.g.*, from approximately amino acids 272-360 and from amino acids 398-421 or from amino acids 398-545; based on *Staphylococcus* numbering in Figure 5B). In the various embodiments, useful fragments comprise those nucleic acid sequences which encode at least the active residue at amino acid position 344 (*Staphylococcus* numbering in Figure 5B) and, preferably, comprise nucleic acid sequences 697-1624, 703-1620, 751-1573, 805-1398, 886-1248, 970-1059, and 997-1044 (*Staphylococcus* numbering in Figures 4A-4C). In other embodiments,

oligonucleotides of microbial GUSes are provided especially for use as amplification primers. In such case, the oligonucleotides are at least 12 bases and preferably at least 15 bases (*e.g.*, at least 18, 21, 25, 30 bases) and generally not longer than 50 bases. It will be appreciated that any of these fragments described herein can be double-stranded, single-stranded, derived from coding strand or complementary strand and be exact or mismatched sequence.

Microbial β-glucuronidase gene products

The present invention also provides β-glucuronidase gene products in various forms. Forms of the GUS protein include, but are not limited to, secreted forms, membrane-bound forms, cytoplasmic forms, fusion proteins, chemical conjugates of GUS and another molecule, portions of GUS protein, and other variants. GUS protein may be produced by recombinant means, biochemical isolation, and the like.

In certain aspects, variants of secreted microbial GUS are useful within the context of this invention. Variants include nucleotide or amino acid substitutions, deletions, insertions, and chimeras (*e.g.*, fusion proteins). Typically, when the result of synthesis, amino acid substitutions are conservative, *i.e.*, substitution of amino acids within groups of polar, non-polar, aromatic, charged, etc. amino acids. As will be appreciated by those skilled in the art, a nucleotide sequence encoding microbial GUS may differ from the wild-type sequence presented in the Figures, due to codon degeneracies, nucleotide polymorphisms, or amino acid differences. In certain embodiments, variants preferably hybridize to the wild-type nucleotide sequence at conditions of normal stringency, which is approximately 25-30°C below Tm of the native duplex (*e.g.*, 1 M Na+ at 65°C; *e.g.* 5X SSPE, 0.5% SDS, 5X Denhardt's solution, at 65°C or equivalent conditions; *see generally*, Sambrook *et al. Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, 1987; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing, 1987). Alternatively, the Tm for other than short oligonucleotides can be calculated by the formula Tm=81.5 + 0.41%(G+C) – log[Na+]. Low stringency hybridizations are performed at conditions

approximately 40°C below Tm, and high stringency hybridizations are performed at conditions approximately 10°C below Tm.

Variants may be constructed by any of the well known methods in the art (*see, generally,* Ausubel *et al.*, *supra*; Sambrook *et al.*, *supra*). Such methods include site-directed oligonucleotide mutagenesis, restriction enzyme digestion and removal or insertion of bases, amplification using primers containing mismatches or additional nucleotides, splicing of another gene sequence to the reference microbial GUS gene, and the like. Briefly, preferred methods for generating a few nucleotide substitutions utilize an oligonucleotide that spans the base or bases to be mutated and contains the mutated base or bases. The oligonucleotide is hybridized to complementary single stranded nucleic acid and second strand synthesis is primed from the oligonucleotide. Similarly, deletions and/or insertions may be constructed by any of a variety of known methods. For example, the gene can be digested with restriction enzymes and religated such that some sequence is deleted or ligated with an isolated fragment having cohesive ends so that an insertion or large substitution is made. In another embodiment, variants are generated by shuffling of regions (see U.S. Patent No. 5,605,793). Variant sequences may also be generated by "molecular evolution" techniques (see U. S. Patent No. 5,723,323). Other means to generate variant sequences may be found, for example, in Sambrook *et al.* (*supra*) and Ausubel *et al.* (*supra*). Verification of variant sequences is typically accomplished by restriction enzyme mapping, sequence analysis, or probe hybridization, although other methods may be used. The double-stranded nucleic acid is transformed into host cells, typically *E. coli*, but alternatively, other prokaryotes, yeast, or larger eukaryotes may be used. Standard screening protocols, such as nucleic acid hybridization, amplification, and DNA sequence analysis, can be used to identify mutant sequences.

In addition to directed mutagenesis in which one or a few amino acids are altered, variants that have multiple substitutions may be generated. The substitutions may be scattered throughout the protein or functional domain or concentrated in a small region. For example, a region may be mutagenized by oligonucleotide-directed mutagenesis in which the oligonucleotide contains a string of

dN bases or the region is excised and replaced by a string of dN bases. Thus, a population of variants with a randomized amino acid sequence in a region is generated. The variant with the desired properties (*e.g.*, more efficient secretion) is then selected from the population.

In preferred embodiments, the protein and variants are capable of being secreted and exhibit β-glucuronidase activity. A GUS protein is secreted if the amount of secretion expressed as a secretion index is statistically significantly higher for the candidate protein compared to a standard, typically *E. coli* GUS. Secretion index maybe calculated as the percentage of total GUS activity in periplasm or other extracellular environment less the percentage of total β-galactosidase activity found in the same extracellular environment.

In other preferred embodiments, a microbial GUS or its variant will exhibit one or more of the biochemical characteristics exhibited by *Staphylococcus* GUS, such as its increased thermal stability, its higher turnover number, and its activity in detergents, presence of end product, high salt conditions and organic solvents as compared to an *E. coli* GUS standard.

In certain preferred embodiments, the microbial GUS is thermostable, having a half-life of at least 10 minutes at 65°C (*e.g.*, at least 14 minutes, 16 minutes, 18 minutes). In other preferred embodiments, GUS protein has a turnover number, expressed as nanomoles of p-nitrophenyl-β-D-glucuronide converted to p-nitrophenol per minute per μg of purified protein, of at least 50 and more preferably at least 60, at least 70, at least 80 and at least 90 nanomoles measured at its temperature optimum. In other preferred embodiments the turnover number is at least 20, at least 30, or at least 40 nanomoles at room temperature. In yet other preferred embodiments, the β -glucuronidase should not be substantially inhibited by the presence of detergents such as SDS (*e.g.*, at 0.1%, 1%, 5%), Triton® X-100 (*e.g.*, at 0.1%, 1%, 5%), or sarcosyl (*e.g.*, at 0.1%, 1%, 5%). In other preferred embodiments, the GUS enzyme is not substantially inhibited (*e.g.*, less than 50% inhibition and more preferably less than 20% inhibition) by either 1 mM or as high as 10 mM glucuronic acid. In still other preferred embodiments, GUS retains substantial activity (at least 50% and preferably at least

70%) in organic solvents, such as dimethylformamide, dimethylsulfoxide and in salt (*e.g.*, NaCl).

In other preferred embodiments, GUS and variants thereof are capable of being secreted and exhibit one or more of the biochemical characteristics disclosed herein. In other embodiments, variants of microbial GUS are capable of binding to a hapten, such as biotin, dinitrophenol, and the like.

In other embodiments, variants may exhibit glucuronide binding activity without enzymatic activity or be directed to other cellular compartments, such as membrane or cytoplasm. Membrane-spanning amino acid sequences are generally hydrophobic and many examples of such sequences are well-known. These sequences may be spliced onto microbial secreted GUS by a variety of methods including conventional recombinant DNA techniques. Similarly, sequences that direct proteins to cytoplasm (*e.g.*, Lys-Asp-Glu-Leu) may be added to the reference GUS, typically by recombinant DNA techniques.

In other embodiments, a fusion protein comprising GUS may be constructed from the nucleic acid molecule encoding microbial and another nucleic acid molecule. As will be appreciated, the fusion partner gene may contribute, within certain embodiments, a coding region. In preferred embodiments, microbial GUS is fused to avidin, streptavidin or an antibody. Thus, it may be desirable to use only the catalytic site of GUS (*e.g.*, amino acids 415-508 reference to *Staphylococcus* sequence). The choice of the fusion partner depends in part upon the desired application. The fusion partner may be used to alter specificity of GUS, provide a reporter function, provide a tag sequence for identification or purification protocols, and the like. The reporter or tag can be any protein that allows convenient and sensitive measurement or facilitates isolation of the gene product and does not interfere with the function of GUS. For example, green fluorescent protein and β-galactosidase are readily available as DNA sequences. A peptide tag is a short sequence, usually derived from a native protein, which is recognized by an antibody or other molecule. Peptide tags include FLAG®, Glu-Glu tag (Chiron Corp., Emeryville, CA), KT3 tag (Chiron Corp.), T7 gene 10 tag (Invitrogen, La Jolla, CA), T7 major capsid protein tag (Novagen, Madison, WI), His_6

(hexa-His), and HSV tag (Novagen). Besides tags, other types of proteins or peptides, such as glutathione-S-transferase may be used.

In other aspects of the present invention, isolated microbial glucuronidase proteins are provided. In one embodiment, GUS protein is expressed as a hexa-His fusion protein and isolated by metal-containing chromatography, such as nickel-coupled beads. Briefly, a sequence encoding His_6 is linked to a DNA sequence encoding a GUS. Although the His_6 sequence can be positioned anywhere in the molecule, preferably it is linked at the 3' end immediately preceding the termination codon. The His-GUS fusion may be constructed by any of a variety of methods. A convenient method is amplification of the GUS gene using a downstream primer that contains the codons for His_6.

In one aspect of the present invention, peptides having microbial GUS sequence are provided. Peptides may be used as immunogens to raise antibodies, as well as other uses. Peptides are generally five to 100 amino acids long, and more usually 10 to 50 amino acids. Peptides are readily chemically synthesized in an automated fashion (*e.g.*, PerkinElmer, ABI Peptide Synthesizer) or may be obtained commercially. Peptides may be further purified by a variety of methods, including high-performance liquid chromatography (HPLC). Furthermore, peptides and proteins may contain amino acids other than the 20 naturally occurring amino acids or may contain derivatives and modification of the amino acids.

β-glucuronidase protein may be isolated by standard methods, such as affinity chromatography using matrices containing saccharose lactone, phenythio- β -glucuronide, antibodies to GUS protein and the like, size exclusion chromatography, ionic exchange chromatography, HPLC, and other known protein isolation methods. (*see generally* Ausubel *et al. supra*; Sambrook *et al. supra*). The protein can be expressed as a hexa-His fusion protein and isolated by metal-affinity chromatography, such as nickel-coupled beads. An isolated purified protein gives a single band on SDS-PAGE when stained with Coomassie brilliant blue.

Antibodies to microbial GUS

Antibodies to microbial GUS proteins, fragments, or peptides discussed herein may readily be prepared. Such antibodies may specifically recognize reference microbial GUS protein and not a mutant (or variant) protein, mutant (or variant) protein and not wild type protein, or equally recognize both the mutant (or variant) and wild-type forms. Antibodies may be used for isolation of the protein, inhibiting (antagonist) activity of the protein, or enhancing (agonist) activity of the protein.

Within the context of the present invention, antibodies are understood to include monoclonal antibodies, polyclonal antibodies, anti-idiotypic antibodies, antibody fragments (*e.g.*, Fab, and $F(ab')_2$, F_v variable regions, or complementarity determining regions). Antibodies are generally accepted as specific against GUS protein if they bind with a K_d of greater than or equal to 10^{-7} M, preferably greater than of equal to 10^{-8} M. The affinity of a monoclonal antibody or binding partner can be readily determined by one of ordinary skill in the art (*see* Scatchard, *Ann. N.Y. Acad. Sci. 51*:660-672, 1949).

Briefly, a polyclonal antibody preparation may be readily generated in a variety of warm-blooded animals such as rabbits, mice, or rats. Typically, an animal is immunized with GUS protein or peptide thereof, which may be conjugated to a carrier protein, such as keyhole limpet hemocyanin. Routes of administration include intraperitoneal, intramuscular, intraocular, or subcutaneous injections, usually in an adjuvant (*e.g.*, Freund's complete or incomplete adjuvant). Particularly preferred polyclonal antisera demonstrate binding in an assay that is at least three times greater than background.

Monoclonal antibodies may also be readily generated from hybridoma cell lines using conventional techniques (*see* U.S. Patent Nos. RE 32,011, 4,902,614, 4,543,439, and 4,411,993; *see also Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Briefly, within one embodiment, a subject animal such as a rat or mouse is injected with GUS or a portion thereof. The protein may be administered as an emulsion in an adjuvant such as Freund's complete or incomplete adjuvant in order to increase the immune response. Between one and three weeks after the initial immunization the animal is generally boosted and may tested for

reactivity to the protein utilizing well-known assays. The spleen and/or lymph nodes are harvested and immortalized. Various immortalization techniques, such as mediated by Epstein-Barr virus or fusion to produce a hybridoma, may be used. In a preferred embodiment, immortalization occurs by fusion with a suitable myeloma cell line (*e.g.*, NS-1 (ATCC No. TIB 18), and P3X63 - Ag 8.653 (ATCC No. CRL 1580) to create a hybridoma that secretes monoclonal antibody. The preferred fusion partners do not express endogenous antibody genes. Following fusion, the cells are cultured in selective medium and are subsequently screened for the presence of antibodies that are reactive against a GUS protein. A wide variety of assays may be utilized, including for example countercurrent immuno-electrophoresis, radioimmunoassays, radioimmunoprecipitations, enzyme-linked immunosorbent assays (ELISA), dot blot assays, western blots, immunoprecipitation, inhibition or competition assays, and sandwich assays (*see* U.S. Patent Nos. 4,376,110 and 4,486,530; *see also Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988).

Other techniques may also be utilized to construct monoclonal antibodies (*see* Huse *et al.*, *Science 246*:1275-1281, 1989; Sastry *et al.*, *Proc. Natl. Acad. Sci. USA 86*:5728-5732, 1989; Alting-Mees *et al.*, *Strategies in Molecular Biology 3*:1-9, 1990; describing recombinant techniques). Briefly, RNA is isolated from a B cell population and utilized to create heavy and light chain immunoglobulin cDNA expression libraries in suitable vectors, such as λImmunoZap(H) and λImmunoZap(L). These vectors may be screened individually or co-expressed to form Fab fragments or antibodies (*see* Huse *et al.*, *supra*; Sastry *et al.*, *supra*). Positive plaques may subsequently be converted to a non-lytic plasmid that allows high level expression of monoclonal antibody fragments from *E. coli*.

Similarly, portions or fragments, such as Fab and Fv fragments, of antibodies may also be constructed utilizing conventional enzymatic digestion or recombinant DNA techniques to yield isolated variable regions of an antibody. Within one embodiment, the genes which encode the variable region from a hybridoma producing a monoclonal antibody of interest are amplified using nucleotide primers for

the variable region, which may be purchased from commercially available sources (*e.g.*, Stratacyte, La Jolla, CA) Amplification products are inserted into vectors such as ImmunoZAP™ H or ImmunoZAP™ L (Stratacyte), which are then introduced into *E. coli*, yeast, or mammalian-based systems for expression. Utilizing these techniques, large amounts of a single-chain protein containing a fusion of the V_H and V_L domains may be produced (*see* Bird *et al.*, *Science 242*:423-426, 1988). In addition, techniques may be utilized to change a "murine" antibody to a "human" antibody, without altering the binding specificity of the antibody.

One of ordinary skill in the art will appreciate that a variety of alternative techniques for generating antibodies exist. In this regard, the following U.S. patents teach a variety of these methodologies and are thus incorporated herein by reference: U.S. Patent Nos. 5,840,479; 5,770,380; 5,204,244; 5,482,856; 5,849,288; 5,780,225; 5,395,750; 5,225,539; 5,110,833; 5,693,762; 5,693,761; 5,693,762; 5,698,435; and 5,328,834.

Once suitable antibodies have been obtained, they may be isolated or purified by many techniques well known to those of ordinary skill in the art (*see Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Suitable techniques include peptide or protein affinity columns, HPLC (*e.g.*, reversed phase, size exclusion, ion-exchange), purification on protein A or protein G columns, or any combination of these techniques.

Assays for function of β-glucuronidase

In preferred embodiments, microbial β-glucuronidase will at least have enzymatic activity and in other preferred embodiments, will also have the capability of being secreted. As noted above, variants of these reference GUS proteins may exhibit altered functional activity and cellular localization. Enzymatic activity may be assessed by an assay such as the ones disclosed herein or in U.S. Patent No. 5,268,463 (Jefferson). Generally, a chromogenic or fluorogenic substrate is incubated with cell extracts, tissue or tissue sections, or purified protein. Cleavage of the substrate is monitored by a method appropriate for the aglycone.

A variety of methods may be used to demonstrate that a β-glucuronidase is secreted. For example, a rapid screening method in which colonies of organisms or cells, such as bacteria, yeast or insect cells, are plated and incubated with a readily visualized glucuronide substrate, such as X-GlcA. A colony with a diffuse staining pattern likely secretes GUS, although such a pattern could indicate that the cell has the ability to pump out the cleaved glucuronide, that the cell has become leaky, or that the enzyme is membrane bound. The unlikely alternatives can be ruled out by using a host cell for transfection that does not pump out cleaved substrate and is deleted for endogenous GUS genes is preferably used.

Secretion of the enzyme may be verified by assaying for GUS activity in the extracellular environment. If the cells secreting GUS are gram-positive bacteria, yeasts, molds, plants, or other organisms with cell walls, activity may be assayed in the culture medium and in a cell extract, however, the protein may not be transported through the cell wall. Thus, if no or low activity of a secreted form of GUS is found in the culture medium, protoplasts made by osmotic shock or enzymatic digestion of the cell wall or other suitable procedure and the supernatant are assayed for GUS activity. If the cells secreting GUS are gram-negative bacteria, culture supernatant is tested, but more likely β-glucuronidase will be retained in the periplasmic space between the inner and outer membrane. In this case, spheroplasts, made by osmotic shock, enzymatic digestion, or other suitable procedure and the supernatant are assayed for GUS activity. Cells without cell walls are assayed for GUS in cell supernatant and cell extracts. The fraction of activity in each compartment is compared to the activity of a non-secreted GUS in the same or similar host cells. A β-glucuronidase is secreted if significantly more enzyme activity than *E. coli* GUS activity is found in extracellular spaces. The amount of secretion is generally normalized to the amount of a non-secreted protein found in extracellular spaces. By this assay, usually less than 10% of *E. coli* GUS is secreted. Within the context of this invention, higher amounts of secreted enzyme are preferred (*e.g.*, greater than 20%, 25%, 30%, 40%, 50%).

β-glucuronidases that exhibit specific substrate specificity are also useful within the context of the present invention. As noted above, glucuronides can be linked

through an oxygen, carbon, nitrogen or sulfur atom. Glucuronide substrates having each of the linkages may be used in one of the assays described herein to identify GUSes that discriminate among the linkages. In addition, various glucuronides containing a variety of aglycones may be used to identify GUSes that discriminate among the aglycones.

Some readily available glucuronides for testing include, but are not limited to:

Phenyl-β-glucuronide
Phenyl β-D-thio-glucuronide
p-Nitrophenyl-β-glucuronide
4-Methylumbelliferyl-β-glucuronide
p-Aminophenyl-β-D-glucuronide
p-Aminophenyl-1-thio-β-D-glucuronide
Chloramphenicol β-D-glucuronide
8-Hydroxyquinoline β-D-glucuronide
5-Bromo-4-chloro-3-indolyl-β-D-glucuronide (X-GlcA)
5-Bromo-6-chloro-3-indolyl-β-D-glucuronide (Magenta-GlcA)
6-Chloro-3-indolyl-β-D-glucuronide (Salmon-β-D-GlcA)
Indoxyl-β-D-glucuronide (Y-GlcA)
Androsterone-3-β-D-glucuronide
α-Naphthyl-β-D-glucuronide
Estriol-3-β-D-glucuronide
17 -β-Estradiol-3-β-D-glucuronide
Estrone-3-β-D-glucuronide
Testosterone-17-β-D-glucuronide
19-nor-Testosterone-17-β-D-glucuronide
Tetrahydrocortisone-3-β-D-glucuronide
Phenolphthalein-β-D-glucuronide
3'-Azido-3'-deoxythymidine-β-D-glucuronide
Methyl-β-D-glucuronide
Morphine-6-β-D-glucuronide

Vectors, host cells and means of expressing and producing protein

Microbial β-glucuronidase may be expressed in a variety of host organisms. For protein production and purification, GUS is preferably secreted and produced in bacteria, such as *E. coli*, for which many expression vectors have been developed and are available. Other suitable host organisms include other bacterial species (*e.g., Bacillus*, and eukaryotes, such as yeast (*e.g., Saccharomyces cerevisiae*),

mammalian cells (*e.g.*, CHO and COS-7), plant cells and insect cells (*e.g.*, Sf9). Vectors for these hosts are well known.

A DNA sequence encoding microbial β-glucuronidase is introduced into an expression vector appropriate for the host. The sequence is derived from an existing clone or synthesized. As described herein, a fragment of the coding region may be used, but if enzyme activity is desired, the catalytic region should be included. A preferred means of synthesis is amplification of the gene from cDNA, genomic DNA, or a recombinant clone using a set of primers that flank the coding region or the desired portion of the protein. Restriction sites are typically incorporated into the primer sequences and are chosen with regard to the cloning site of the vector. If necessary, translational initiation and termination codons can be engineered into the primer sequences. The sequence of GUS can be codon-optimized for expression in a particular host. For example, a secreted form of β-glucuronidase isolated from a bacterial species that is expressed in a fungal host, such as yeast, can be altered in nucleotide sequence to use codons preferred in yeast. Codon-optimization may be accomplished by methods such as splice overlap extension, site-directed mutagenesis, automated synthesis, and the like.

At minimum, an expression vector must contain a promoter sequence Other regulatory sequences may be included. Such sequences include a transcription termination signal sequence, secretion signal sequence, origin of replication, selectable marker, and the like. The regulatory sequences are operationally associated with one another to allow transcription or translation.

Expression in bacteria

The plasmids used herein for expression of secreted GUS include a promoter designed for expression of the proteins in a bacterial host. Suitable promoters are widely available and are well known in the art. Inducible or constitutive promoters are preferred. Such promoters for expression in bacteria include promoters from the T7 phage and other phages, such as T3, T5, and SP6, and the trp, lpp, and lac operons. Hybrid promoters (*see*, U.S. Patent No. 4,551,433), such as tac and trc, may also be used. Promoters for expression in eukaryotic cells include the P10 or polyhedron gene

promoter of baculovirus/insect cell expression systems (*see, e.g.*, U.S. Patent Nos. 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784), MMTV LTR, RSV LTR, SV40, metallothionein promoter (*see, e.g.*, U.S. Patent No. 4,870,009) and other inducible promoters. For protein expression, a promoter is inserted in operative linkage with the coding region for β-glucuronidase.

The promoter controlling transcription of β-glucuronidase may be controlled by a repressor. In some systems, the promoter can be derepressed by altering the physiological conditions of the cell, for example, by the addition of a molecule that competitively binds the repressor, or by altering the temperature of the growth media. Preferred repressor proteins include, but are not limited to the *E. coli* lacI repressor responsive to IPTG induction, the temperature sensitive λcI857 repressor, and the like. The *E. coli* lacI repressor is preferred.

In other preferred embodiments, the vector also includes a transcription terminator sequence. A "transcription terminator region" has either a sequence that provides a signal that terminates transcription by the polymerase that recognizes the selected promoter and/or a signal sequence for polyadenylation.

Preferably, the vector is capable of replication in host cells. Thus, for bacterial hosts, the vector preferably contains a bacterial origin of replication. Preferred bacterial origins of replication include the f1-ori and col E1 origins of replication, especially the origin derived from pUC plasmids.

The plasmids also preferably include at least one selectable gene that is functional in the host. A selectable gene includes any gene that confers a phenotype on the host that allows transformed cells to be identified and selectively grown. Suitable selectable marker genes for bacterial hosts include the ampicillin resistance gene (Amp^r), tetracycline resistance gene (Tc^r) and kanamycin resistance gene (Kan^r). Suitable markers for eukaryotes usually complement a deficiency in the host (*e.g.*, thymidine kinase (tk) in tk- hosts). However, drug markers are also available (*e.g.*, G418 resistance and hygromycin resistance).

The sequence of nucleotides encoding β-glucuronidase may also include a classical secretion signal, whereby the resulting peptide is a precursor protein

processed and secreted. The resulting processed protein may be recovered from the periplasmic space or the fermentation medium. Secretion signals suitable for use are widely available and are well known in the art (von Heijne, *J. Mol. Biol. 184*:99-105, 1985). Prokaryotic and eukaryotic secretion signals that are functional in *E. coli* (or other host) may be employed. The presently preferred secretion signals include, but are not limited to pelB, matα, extensin and glycine-rich protein.

One skilled in the art appreciates that there are a wide variety of suitable vectors for expression in bacterial cells and which are readily obtainable. Vectors such as the pET series (Novagen, Madison, WI) and the tac and trc series (Pharmacia, Uppsala, Sweden) are suitable for expression of a β-glucuronidase. A suitable plasmid is ampicillin resistant, has a colEI origin of replication, $lacI^q$ gene, a lac/trp hybrid promoter in front of the lac Shine-Dalgarno sequence, a hexa-his coding sequence that joins to the 3' end of the inserted gene, and an rrnB terminator sequence.

The choice of a bacterial host for the expression of a β-glucuronidase is dictated in part by the vector. Commercially available vectors are paired with suitable hosts. The vector is introduced in bacterial cells by standard methodology. Typically, bacterial cells are treated to allow uptake of DNA (for protocols, see generally, Ausubel *et al.*, *supra*; Sambrook *et al.*, *supra*). Alternatively, the vector may be introduced by electroporation, phage infection, or another suitable method.

Expression in plant cells

As noted above, the present invention provides vectors capable of expressing microbial secreted β-glucuronidase and secreted microbial β-glucuronidases. For agricultural applications, the vectors should be functional in plant cells. Suitable plants include, but are not limited to, wheat, rice, corn, soybeans, lupins, vegetables, potatoes, canola, nut trees, coffee, cassava, yam, alfalfa and other forage plants, cereals, legumes and the like. In one embodiment, rice is a host for GUS gene expression.

Vectors that are functional in plants are preferably binary plasmids derived from *Agrobacterium* plasmids. Such vectors are capable of transforming plant cells. These vectors contain left and right border sequences that are required for

integration into the host (plant) chromosome. At minimum, between these border sequences is the gene to be expressed under control of a promoter. In preferred embodiments, a selectable gene is also included. The vector also preferably contains a bacterial origin of replication for propagation in bacteria.

A gene for microbial β-glucuronidase should be in operative linkage with a promoter that is functional in a plant cell. Typically, the promoter is derived from a host plant gene, but promoters from other plant species and other organisms, such as insects, fungi, viruses, mammals, and the like, may also be suitable, and at times preferred. The promoter may be constitutive or inducible, or may be active in a certain tissue or tissues (tissue type-specific promoter), in a certain cell or cells (cell-type specific promoter), of at a particular stage or stages of development (development-type specific promoter). The choice of a promoter depends at least in part upon the application. Many promoters have been identified and isolated (*e.g.*, CAMV35S promoter, maize Ubiquitin promoter) (*see, generally*, GenBank and EMBL databases). Other promoters may be isolated by well-known methods. For example, a genomic clone for a particular gene can be isolated by probe hybridization. The coding region is mapped by restriction mapping, DNA sequence analysis, RNase probe protection, or other suitable method. The genomic region immediately upstream of the coding region comprises a promoter region and is isolated. Generally, the promoter region is located in the first 200 bases upstream, but may extend to 500 or more bases. The candidate region is inserted in a suitable vector in operative linkage with a reporter gene, such as in pBI121 in place of the CaMV 35S promoter, and the promoter is tested by assaying for the reporter gene after transformation into a plant cell. (*see, generally,* Ausubel *et al.*, *supra*; Sambrook *et al.*, *supra*; *Methods in Plant Molecular Biology and Biotechnology*, Ed. Glick and Thompson, CRC Press, 1993.)

Preferably, the vector contains a selectable marker for identifying transformants. The selectable marker preferably confers a growth advantage under appropriate conditions. Generally, selectable markers are drug resistance genes, such as neomycin phosphotransferase. Other drug resistance genes are known to those in the art and may be readily substituted. Selectable markers include, ampicillin resistance,

tetracycline resistance, kanamycin resistance, chloramphenicol resistance, and the like. The selectable marker also preferably has a linked constitutive or inducible promoter and a termination sequence, including a polyadenylation signal sequence. Other selection systems, such as positive selection can alternatively be used (U.S. Patent Nos._____).

The sequence of nucleotides encoding β-glucuronidase may also include a classical secretion signal, whereby the resulting peptide is a precursor protein processed and secreted. Suitable signal sequences of plant genes include, but are not limited to the signal sequences from glycine-rich protein and extensin. In addition, a glucuronide permease gene to facilitate uptake of glucuronides may be co-transfected either from the same vector containing microbial GUS or from a separate expression vector.

A general vector suitable for use in the present invention is based on pBI121 (U.S. Patent No. 5,432,081) a derivative of pBIN19. Other vectors have been described (U.S. Patent Nos. 4,536,475; 5,733,744; 4,940,838; 5,464,763; 5,501,967; 5,731,179) or may be constructed based on the guidelines presented herein. The plasmid pBI121 contains a left and right border sequence for integration into a plant host chromosome and also contains a bacterial origin of replication and selectable marker. These border sequences flank two genes. One is a kanamycin resistance gene (neomycin phosphotransferase) driven by a nopaline synthase promoter and using a nopaline synthase polyadenylation site. The second is the *E. coli* GUS gene (reporter gene) under control of the CaMV 35S promoter and polyadenlyated using a nopaline synthase polyadenylation site. The *E. coli* GUS gene is replaced with a gene encoding a secreted form of β-glucuronidase. If appropriate, the CaMV 35S promoter is replaced by a different promoter. Either one of the expression units described above is additionally inserted or is inserted in place of the CaMV promoter and GUS gene.

Plants may be transformed by any of several methods. For example, plasmid DNA may be introduced by *Agrobacterium* co-cultivation (*e.g.*, U.S. Patent No. 5,591,616; 4,940,838) or bombardment (*e.g.*, U.S. Patent No. 4,945,050; 5,036,006; 5,100,792; 5,371,015). Other transformation methods include electroporation (U.S.

Patent No. 5,629,183), $CaPO_4$-mediated transfection, gene transfer to protoplasts (AU B 600221), microinjection, and the like (*see, Gene Transfer to Plants*, Ed. Potrykus and Spangenberg, Springer, 1995, for procedures). Preferably, vector DNA is first transfected into *Agrobacterium* and subsequently introduced into plant cells. Most preferably, the infection is achieved by *Agrobacterium* co-cultivation. In part, the choice of transformation methods depends upon the plant to be transformed. Tissues can alternatively be efficiently infected by *Agrobacterium* utilizing a projectile or bombardment method. Projectile methods are generally used for transforming sunflowers and soybean. Bombardment is often used when naked DNA, typically *Agrobacterium* binary plasmids or pUC-based plasmids, is used for transformation or transient expression.

Briefly, co-cultivation is performed by first transforming *Agrobacterium* by freeze-thaw method (Holsters *et al.*, *Mol. Gen. Genet. 163*: 181-187, 1978) or by other suitable methods (*see,* Ausubel, *et al. supra*; Sambrook *et al.*, *supra*). Briefly, a culture of *Agrobacterium* containing the plasmid is incubated with leaf disks, protoplasts, meristematic tissue, or calli to generate transformed plants (Bevan, *Nucl. Acids. Res. 12*:8711, 1984) (U.S. Patent No. 5,591,616). After co-cultivation for about 2 days, bacteria are removed by washing and plant cells are transferred to plates containing antibiotic (*e.g.*, cefotaxime) and selecting medium. Plant cells are further incubated for several days. The presence of the transgene may be tested for at this time. After further incubation for several weeks in selecting medium, calli or plant cells are transferred to regeneration medium and placed in the light. Shoots are transferred to rooting medium and then into glass house.

Briefly, for microprojectile bombardment, cotyledons are broken off to produce a clean fracture at the plane of the embryonic axis, which are placed cut surface up on medium with growth regulating hormones, minerals and vitamin additives. Explants from other tissues or methods of preparation may alternatively be used. Explants are bombarded with gold or tungsten microprojectiles by a particle acceleration device and cultured for several days in a suspension of transformed *Agrobacterium*. Explants are transferred to medium lacking growth regulators but

containing drug for selection and grown for 2-5 weeks. After 1-2 weeks more without drug selection, leaf samples from green, drug-resistant shoots are grafted to in vitro grown rootstock and transferred to soil.

A positive selection system, such as using cellobiuronic acid and culture medium lacking a carbon source, is preferably used (*see*, co-pending application no. 09/130,695).

Activity of secreted GUS is conveniently assayed in whole plants or in selected tissues using a glucuronide substrate that is readily detected upon cleavage. Glucuronide substrates that are colorimetric are preferred. Field testing of plants may be performed by spraying a plant with the glucuronide substrate and observing color formation of the cleaved product.

Classical tests for a transgene such as Southern blotting and hybridization or genetic segregation can also be performed.

Expression in other organisms

A variety of other organisms are suitable for use in the present invention. For example, various fungi, including yeasts, molds, and mushrooms, insects, especially vectors for diseases and pathogens, and other animals, such as cows, mice, goats, birds, aquatic animals (*e.g.*, shrimp, turtles, fish, lobster and other crustaceans), amphibians and reptiles and the like, may be transformed with a GUS transgene.

The principles that guide vector construction for bacteria and plants, as discussed above, are applicable to vectors for these organisms. In general, vectors are well known and readily available. Briefly, the vector should have at least a promoter functional in the host in operative linkage with GUS. Usually, the vector will also have one or more selectable markers, an origin of replication, a polyadenylation signal and transcription terminator.

The sequence of nucleotides encoding β-glucuronidase may also include a classical secretion signal, whereby the resulting peptide is a precursor protein processed and secreted. Suitable secretion signals may be obtained from a variety of genes, such as mat-alpha or invertase genes. In addition, a permease gene may be co-transfected.

One of ordinary skill in the art will appreciate that a variety of techniques for producing transgenic animals exist. In this regard, the following U.S. patents teach such methodologies and are thus incorporated herein by reference: U.S. Patent Nos. 5,162,215; 5,545,808; 5,741,957; 4,873,191; 5,780,009; 4,736,866; 5,567,607; and 5,633,076.

Uses of microbial β-glucuronidase

As noted above, microbial β-glucuronidase may be used in a variety of applications. In certain aspects, microbial β-glucuronidase can be used as a reporter/effector molecule and as a diagnostic tool. As taught herein, microbial β-glucuronidase that is secretable is preferred as an *in vivo* reporter/effector molecule, whereas, in *in vitro* diagnostic applications, the biochemical characteristics of the β-glucuronidase disclosed herein (*e.g.*, thermal stability, high turnover number) may provide preferred advantages.

Microbial GUS, either secreted or non-secreted, can be used as a marker/effector for transgenic constructions. In a certain embodiments, the transgenic host is a plant, such as rice, corn, wheat, or an aquatic animal. The transgenic GUS may be used in at least three ways: one in a method of positive selection, obviating the need for drug resistance selection, a second as a system to target molecules to specific cells, and a third as a means of detecting and tracking linked genes.

For positive selection, a host cell, (*e.g.*, plant cells) is transformed with a GUS (preferably secretable GUS) transgene. Selection is achieved by providing the cells with a glucuronidated form of a required nutrient (U.S. Patent Nos 5,994,629; 5,767,378; PCT US99/17804). For example, all cells require a carbon source, such as glucose. In one embodiment, glucose is provided as glucuronyl glucose (cellobiuronic acid), which is cleaved by GUS into glucose plus glucuronic acid. The glucose would then bind to receptors and be taken up by cells. The glucuronide can be any required compound, including without limitation, a cytokinin, auxin, vitamin, carbohydrate, nitrogen-containing compound, and the like. It will be appreciated that this positive selection method can be used for cells and tissues derived from diverse organisms, such

as animal cells, insect cells, fungi, and the like. The choice of glucuronide will depend in part upon the requirements of the host cell..

As a marker/effector molecule, secreted GUS (s-GUS) is preferred because it is non-destructive, that is, the host does not need to be destroyed in order to assay enzyme activity. A non-destructive marker has special utility as a tool in plant breeding. The GUS enzyme can be used to detect and track linked endogenous or exogenously introduced genes. GUS may also be used to generate sentinel plants that serve as bioindicators of environmental status. Plant pathogen invasion can be monitored if GUS is under control of a pathogen promoter. In addition, such transgenic plants may serve as a model system for screening inhibitors of pathogen invasion. In this system, GUS is expressed if a pathogen invades. In the presence of an effective inhibitor, GUS activity will not be detectable. In certain embodiments, GUS is co-transfected with a gene encoding a glucuronide permease.

Preferred transgenes for introduction into plants encode proteins that affect fertility, including male sterility, female fecundity, and apomixis; plant protection genes, including proteins that confer resistance to diseases, bacteria, fungus, nematodes, viruses and insects; genes and proteins that affect developmental processes or confer new phenotypes, such as genes that control meristem development, timing of flowering, cell division or senescence (*e.g.*, telomerase) toxicity (*e.g.*, diphtheria toxin, saporin) affect membrane permeability (*e.g.*, glucuronide permease (U.S. Patent No. 5,432,081)), transcriptional activators or repressors, and the like.

Insect and disease resistance genes are well known. Some of these genes are present in the genome of plants and have been genetically identified. Others of these genes have been found in bacteria and are used to confer resistance.

Particularly well known insect resistance genes are the crystal genes of *Staphylococcus thuringiensis*. The crystal genes are active against various insects, such as lepidopterans, *Diptera*, *Hemiptera* and *Coleoptera*. Many of these genes have been cloned. For examples, see, GenBank; U.S. Patent Nos. 5,317,096; 5,254,799; 5,460,963; 5,308,760, 5,466,597, 5,2187,091, 5,382,429, 5,164,180, 5,206,166, 5,407,825, 4,918,066. Gene sequences for these and related proteins may be obtained

by standard and routine technologies, such as probe hybridization of a *B. thuringiensis* library or amplification (*see generally*, Sambrook *et al.*, *supra*, Ausubel *et al. supra*). The probes and primers may be synthesized based on publicly available sequence information.

Other resistance genes to *Sclerotinia*, cyst nematodes, tobacco mosaic virus, flax and crown rust, rice blast, powdery mildew, verticillum wilt, potato beetle, aphids, as well as other infections, are useful within the context of this invention. Examples of such disease resistance genes may be isolated from teachings in the following references: isolation of rust disease resistance gene from flax plants (WO 95/29238); isolation of the gene encoding Rps2 protein from *Arabidopsis thaliana* that confers disease resistance to pathogens carrying the avrRpt2 avirulence gene (WO 95/28478); isolation of a gene encoding a lectin-like protein of kidney bean confers insect resistance (JP 71-32092); isolation of the Hm1 disease resistance gene to *C. carbonum* from maize (WO 95/07989); for examples of other resistance genes, see WO 95/05743; U.S. Patent No. 5,496,732; U.S. Patent No. 5,349,126; EP 616035; EP 392225; WO 94/18335; JP 43-20631; EP 502719; WO 90/11770; U.S. Patent 5,270,200; U.S. Patent Nos. 5,218,104 and 5,306,863). In addition, general methods for identification and isolation of plant disease resistance genes are disclosed (WO 95/28423). Any of these gene sequences suitable for insertion in a vector according to the present invention may be obtained by standard recombinant technology techniques, such as probe hybridization or amplification. When amplification is performed, restriction sites suitable for cloning are preferably inserted. Nucleotide sequences for other transgenes, such as controlling male fertility, are found in U.S. Patent No. 5,478,369, references therein, and Mariani *et al.*, *Nature 347*:737, 1990.

In similar fashion, microbial GUS, preferably secreted, can be used to generate transgenic insects for tracking insect populations or facilitate the development of a bioassay for compounds that affect molecules critical for insect development (*e.g.*, juvenile hormone). Secreted GUS may also serve as a marker for beneficial fungi destined for release into the environment. The non-destructive marker is useful for detecting persistence and competitive advantage of the released organisms.

In animal systems, secreted GUS may be used to achieve extracellular detoxification of glucuronides (*e.g,* toxin glucuronide) and examine conjugation patterns of glucuronides. Furthermore, as discussed above, secreted GUS may be used as a transgenic marker to track cells or as a positive selection system, or to assist in development of new bioactive GUS substrates that do not need to be transported across membrane. Aquatic animals are suitable hosts for GUS transgene. GUS may be used in these animals as a marker or effector molecule.

Within the context of this invention, GUS may also be used in a system to target molecules to cells. This system is particularly useful when the molecules are hydrophobic and thus, not readily delivered. These molecules can be useful as effectors (*e.g.,* inducers) of responsive promoters. For example, molecules such as ecdysone are hydrophobic and not readily transported through phloem in plants. When ecdysone is glucuronidated it becomes amphipathic and can be delivered to cells by way of phloem. Targeting of compounds such as ecdysone-glucuronic acid to cells is accomplished by causing cells to express receptor for ecdysone. As ecdysone receptor is naturally only expressed in insect cells, however a host cell that is transgenic for ecdysone receptor will express it. The glucuronide containing ecdysone then binds only to cells expressing the receptor. If these cells also express GUS, ecdysone will be released from the glucuronide and able to induce expression from an ecdysone-responsive promoter. Plasmids containing ecdysone receptor genes and ecdysone responsive promoter can be obtained from Invitrogen (Carlsbad, CA). Other ligand-receptors suitable for use in this system include glucocorticoids/glucocorticoid receptor, estrogen/estrogen receptor, antibody and antigen, and the like (*see* also U.S. Patent Nos. 5,693,769 and 5,612,317).

In another aspect, purified microbial β-glucuronidase is used in medical applications. For these applications, secretion is not a necessary characteristic although it may be a desirable characteristic for production and purification. The biochemical attributes, such as the increased stability and enzymatic activity disclosed herein are preferred characteristics. The microbial glucuronidase preferably has one or more of the disclosed characteristics.

For the majority of drug or pharmaceutical analysis, the compounds in urine, blood, saliva, or other bodily fluids are de-glucuronidated prior to analysis. Such a procedure is undertaken because compounds are often, if not nearly always, detoxified by glucuronidation in vertebrates. Thus, drugs that are in circulation and have passed through a site of glucuronidation (*e.g.*, liver) are found conjugated to glucuronic acid. Such glucuronides yield a complex pattern upon analysis by, for example, HPLC. However, after the aglycone (drug) is cleaved from the glucuronic acid, a spectrum can be compared to a reference spectrum. Currently, *E. coli* GUS is utilized in medical diagnostics, but as shown herein, microbial GUS, *e.g. Staphylococcus* GUS has superior qualities.

The microbial GUS enzymes disclosed herein may be used in traditional medical diagnostic assays, such as described above for drug testing, pharmacokinetic studies, bioavailability studies, diagnosis of diseases and syndromes, following progression of disease or its response to therapy and the like (*see* U.S. Patent Nos. 5,854,009, 4,450,239, 4,274,832, 4,473,640, 5,726,031, 4,939,264, 4,115,064, 4,892,833). These β-glucuronidase enzymes may be used in place of other traditional enzymes (*e.g.*, alkaline phosphatase, horseradish peroxidase, beta-galactosidase, and the like) and compounds (*e.g.*, green fluorescent protein, radionuclides) that serve as visualizing agents. Microbial GUS has qualities advantageous for use as a visualizing agent: it is highly specific for the substrate, water soluble and the substrates are stable. Thus, microbial GUS is suitable for use in Southern analysis of DNA, Northern analysis, ELISA, and the like.

In preferred embodiments, microbial GUS binds a hapten, either as a fusion protein with a partner protein that binds the hapten (*e.g.*, avidin that binds biotin, antibody) or alone. If used alone, microbial GUS can be mutagenized and selected for hapten-binding abilities. Mutagenesis and binding assays are well known in the art. In addition, microbial GUS can be conjugated to avidin, streptavidin, antibody or other hapten binding protein and used as a reporter in the myriad assays that currently employ enzyme-linked binding proteins. Such assays include immunoassays, Western blots, *in situ* hybridizations, HPLC, high-throughput binding assays, and the like (see, for

examples, U.S. Patent Nos. 5,328,985 and 4,839,293, which teach avidin and streptavidin fusion proteins and U.S. Patent No. 4,298,685, Diamandis and Christopoulos, *Clin. Chem. 37*:625, 1991; Richards, *Methods Enzymol. 184*:3, 1990; Wilchek and Bayer, *Methods Enzymol. 184*:467, 1990; Wilchek and Bayer, *Methods Enzymol. 184*:5, 1990; Wilchek and Bayer, *Methods Enzymol. 184*:14, 1990; Dunn, *Methods Mol. Biol. 32*:227, 1994; Bloch, *J. Hitochem. Cytochem. 41*:1751, 1993; Bayer and Wilchek *J. Chromatogr. 510*:3, 1990, which teach various applications of enzyme-linked technologies and methods).

Microbial GUSes can also be used in therapeutic methods. By glucuronidating compounds such as drugs, the compound is inactivated. When a glucuronidase is expressed or targeted to the site for delivery, the glucuronide is cleaved and the compound delivered. For these purposes, GUS may be expressed as a transgene or delivered, for example, coupled to an antibody specific for the target cell (*see e.g.*, U.S. Patent Nos. 5,075,340, 4,584,368, 4,481,195, 4,478,936. 5,760,008, 5,639,737, 4,588,686).

The present invention also provides kits comprising microbial GUS protein or expression vectors containing microbial GUS gene. One exemplary type of kit is a dipstick test. Such tests are widely utilized for establishing pregnancy, as well as other conditions. Generally, these dipstick tests assay the glucuronide form, but it would be advantageous to use reagents that detect the aglycone form. Thus, GUS may be immobilized on the dipstick adjacent to or mixed in with the detector molecule (*e.g.*, antibody). The dipstick is then dipped in the test fluid (*e.g.*, urine) and as the compounds flow past GUS, they are cleaved into aglycone and glucuronic acid. The aglycone is then detected. Such a setup may be extremely useful for testing compounds that are not readily detectable as glucuronides.

In a variation of this method, the microbial GUS enzyme is engineered to bind a glucuronide, but lack enzymatic activity. The enzyme will then bind the glucuronide and the enzyme is detected by standard methodology. Alternatively, GUS is fused to a second protein, either as a fusion protein or as a chemical conjugate, that binds an aglycone. The fusion is incubated with the test substance and an indicator

substrate is added. This procedure may be used for ELISA, Northern, Southern analysis and the like.

The following examples are offered by way of illustration, and not by way of limitation.

EXAMPLES

EXAMPLE 1

IDENTIFICATION OF MICROBES THAT EXPRESS β-GLUCURONIDASE

Skin microbes are obtained using cotton swabs immersed in 0.1% Triton® X-100 and rubbing individual arm pits or by dripping the solution directly into arm pits and recovering it with a pipette. Seven individuals are sampled. Dilutions (1:100, 1:1000) of arm pit swabs are plated on 0.1X and 0.5X TSB (Tryptone Soy Broth, Difco) agar containing 50 μg/mL X-GlcA (5-bromo-4-chloro-3-indolyl β-D-glucuronide), an indicator substrate for β-glucuronidase. This substrate gives a blue precipitate at the site of enzyme activity (see U.S. Patent No. 5,268,463). TSB is a rich medium which promotes growth of a wide range of microorganisms. Plates are incubated at 37°C.

Soil samples (ca. 1 g) are obtained from an area in Canberra, ACT, Australia (10 samples) and from Queanbeyan, NSW, Australia (12 samples). Although only one of the ten samples from Canberra is intentionally taken from an area of pigeon excrement, most isolates displaying β-glucuronidase activity are in the genera *Enterobacter* or *Salmonella*. Soil samples are shaken in 1-2 mL of water; dilutions of the supernatant are treated as for skin samples, except that incubation is at 30°C and 1.0X TSB plates are used rather than diluted TSB. Some bacteria lose vitality if maintained on diluted medium, although the use of full-strength TSB usually delays, but does not prevent, the onset of indigo build up from X-GlcA hydrolysis.

Microbes that secrete β-glucuronidase have a strong, diffuse staining pattern (halo) surrounding the colony. The appearance of blue colonies varies in time, from one to several days. Under these conditions (aerobic atmosphere and rich medium) many microorganisms grow. Of these, approximately 0.1-1% display β-glucuronidase phenotype, with the secretory phenotype being less common than the non-secretory phenotype.

Colonies that exhibit a strong, diffuse staining pattern are selected for further purification, which consists of two or more streaking of those colonies.

Occasionally segregation of color production can be observed after the purification procedure. In Table 1 below, a summary of the findings is presented. Some strains are listed as GUS secretion-negative because a later repetition of the halo test was negative, showing that the phenotype can vary, possibly because of growth conditions.

Phylogenetic analysis

For phylogenetic identification of the microbes, a variable region of 16S rDNA is amplified using primers, P3-16SrDNA and 1100r-16SrDNA (*see* Table 2), derived from two conserved regions within stem-loop structures of the rRNA. The amplified region corresponds to nucleotides 361 to 705 of *E. coli* rRNA, including the primers. Amplification conditions for 16S rDNA are 94°C for 2 min; followed by 35 cycles of 94°C for 20 sec, 48°C for 40 sec, 72°C for 1.5 min; followed by incubation at 72°C for 5 min.

Amplified fragments are separated by electrophoresis on TAE agarose gels (approximately 1.2%), excised and extracted by freeze-fracture and phenol treatment. Fragments are further purified using Qiagen (Clifton Hill, Vic, Australia) silica-based membranes in microcentrifuge tubes. Purified DNA fragments are sequenced using the amplification primers in combination with BigDye™ Primer Cycle Sequencing Kit from Perkin-Elmer ABI (fluorescent dye termal cycling sequencing) (Foster City, CA). Cycling conditions for DNA sequence reactions are: 2 min at 94°C, followed by 30 cycles of 94°C for 30 sec, 50°C for 15 sec, and 60°C for 2 min. A 10μL reaction uses 4 μL of BigDye™ Terminator mix, 1 μL of 10 μM primer, and 200-500 ng of DNA. The reaction products are precipitated with ethanol or iso-propanol, resuspended and subjected to gel separation and nucleotide analysis.

The ribosomal sequences are aligned and assigned to phylogenetic placement using the facilities of the Ribosomal Database Project of Michigan State University (rdpwww.life.uiuc.edu which now contains more than 10,000 16S rRNA sequences (Maidak *et al., Nucl. Acids Res. 27*:171-173; 1999). Phylogenetic placement is used to select strains for further study.

Table 1

STRAIN	GUS Secretion	GUS Amplif	Genus and tentative species	Phylogenetic position
SKIN				
EH2	+	yes	Staphylococcus warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
EH4	+	yes	Staphylococcus warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
EH4-110A	-	yes	Staphylococcus warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
LS-B	+	yes	Staphylococcus haemophilus/hominis	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
PG3A	+	no	Staphylococcus hominis/warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
SH1B	+	no	Staphylococcus warneri/aureus	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
SH1C	+	yes	Staphylococcus warneri/aureus	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
CRA1	+	no	Staphylococcus warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
CRA2	+	no	Staphylococcus warneri	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
CANBERRA SOIL				
CSW1a	-	yes	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
CSW1b		yes	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
CDS1	+	no	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
CBP1	-	yes	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
CS2.1	-	no	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
CS2.3	-	no	Salmonella/Enterobacter	Proteobacteria - Gamma Subdivision - Enterics and Relatives
QUEANBEYAN SOIL				
Q1.2	-	yes	Pseudomonas/Azospirillum	Proteobacteria - Gamma Subdivision - Pseudomonas and Relatives
Q1.3	+	no	Arthrobacter	Firmicutes - Actinobacteria - Micrococcineae
Q2VD3	-	yes	Pseudomonas/Azospirillum	Proteobacteria - Gamma Subdivision - Pseudomonas and Relatives
Q2VD6	-	yes	Arthrobacter	Firmicutes - Actinobacteria - Micrococcineae
Q2VD7	-	yes	Clavibacterium	Firmicutes - Actinobacteria - Micrococcineae
Q3WR2	+	no	Planococcus	Firmicutes / Bacillus-Lactobacillus-Streptococcus Subdivision
Q3WR6	+	yes	Micrococcus	Firmicutes - Actinobacteria - Micrococcineae
Q4DS1	-	no	Curtobacterium	Firmicutes - Actinobacteria - Micrococcineae
QRM1	-	no	Arthrobacter	Firmicutes - Actinobacteria - Micrococcineae
QRM2	-	no	Arthrobacter	Firmicutes - Actinobacteria - Micrococcineae

QRM6	-	no	Pseudomonas	Proteobacteria - Gamma Subdivision - Pseudomonas and Relatives
QTCR3	+	no	Arthrobacter	Firmicutes - Actinobacteria - Micrococcineae

^ where two genera or species are listed, the rRNA analysis is inconclusive

As can be observed from the table above, all GUS expressing skin isolates belong to the genus *Staphylococcus* and to a limited number of species, *Staphylococcus warneri* and *Staphylococcus homini* or *haemophilus*. The Canberra soil samples all belonged to the genera *Salmonella/Enterobacter* (bacteria are herein referred to in shorthand as *Salmonella*). These two genera are very similar in the 16S rRNA, thus a conclusive identification of the genus requires additional analyses. In contrast, a higher degree of microbial diversity was found in the Queanbeyan strains. Several bacteria are chosen for further studies.

The presence of GUS genes is established by amplification using degenerate oligonucleotides derived from a conserved region of the GUS gene. A pair of oligonucleotides is designed using an alignment of *E. coli* gusA and human GUS sequences. The primer T3-GUS-2F covers *E. coli* GUS amino acids 163-168 (DFFNYA), while T7-GUS-5B covers the complementary sequence to amino acids 549-553 (WNFAD). The full length of *E. coli* GUS is 603 amino acids. As shown in Table 1, amplification is not always successful, likely due to mismatching of the primers with template. Thus, a negative amplification does not necessarily signify that the microorganism lacks a GUS gene.

EXAMPLE 2

CLONING OF GUS GENES BY GENETIC COMPLEMENTATION

Genomic DNA of several candidate strains is isolated and digested with one of the following enzymes, *Eco*R I, *Bam*H I, *Hind* III, *Pst* I. Digested DNA fragments are ligated into the corresponding site of plasmid vector pBluescript II SK (+), and the ligation mix is electroporated into *E. coli* KW1, which is a strain deleted for the complete GUS operon. Colonies are plated on LB-X-GlcA plates and assayed

for blue color. Halo formation is not used as a criterium, because behavior of the GUS gene in a different genetic background may alter the phenotype or detectability. In general though, halo formation is obtained in KW1.

Isolated plasmids from GUS+ transformants are retransformed into KW1 and also into DH5α to demonstrate that the GUS gene is contained within the construct. In all cases, retransformant colonies stained blue with X-GlcA.

EXAMPLE 3

DNA SEQUENCE ANALYSIS OF GUS GENES ISOLATED BY COMPLEMENTATION

DNA sequence is determined for the isolates that amplified from the primers T3 and T7, which flank the pBS polylinker. Cyclic thermal sequencing was done as above, except that elongation time is increased to 4 min to allow for longer sequence determinations. Alternatively, transposon mutagenesis was used to introduce sequencing primer sites randomly into the GUS gene (GPS kit; New England Biolabs, MA, USA).

The sequence information is used to design new oligonucleotides to obtain the full-length sequence of the clones.

Table 2

PRIMER	BASES	Tm	SEQUENCE	SEQ ID No
GUS-2T	16	30.3	AYT TYT TYA AYT AYG C	
GUS-5B	18	49.5	GAA RTC IGC RAA RTT CCA	
CSW-RTSHY(F)	17	47.9	ATC GCA CGT CCC ACT AC	
CSW-RTSHY(R)	18	47.9	CGT GCG ATA GGA GTT AGC	
EH-FRTSHY(F)	22	46.1	ATT TAG AAC ATC TCA TTA TCC C	
EH-FRTSHY(R)	23	47.6	TGA GAT GTT CTA AAT GAA TTA GC	
LSB-KRPVT(R)	17	53.2	ATC GTG ACC GGA CGC TT	
CBP-QAYDE	17	51.1	GCG CGT AAT CTT CCT GG	
NG-RP1L	18	59.7	TAG C(GA)C CTT CGC TTT CGG	
NG-RP1R	20	40.7	ATC ATG TTT ACA GAG TAT GG	
Tm-MVRPQRN	17	48.4	ATG GTA AGA CCG CAA CG	
Tm-Nco-MVRPQRN	25	61.8	TAA AAA CCA TGG TAA GAC CGC AAC G	___

PRIMER	BASES	Tm	SEQUENCE	SEQ ID No
Tm-RRLWSE(R)	20	47.9	CCT CAC TCC ACA GTC TTC TC	
Tm-RRLWSE(R)-Nhe	30	67.4	AGA CCG CTA GCC TCA CTC CAC AGT CTT CTC	
Ps-FDFFNYA(F)	22	47.1	TTT GAC TTT TTC AAC TAT GCA G	
Ps-DFFNYA(R)	23	47.2	AAT TCT GCA TAG TTG AAA AAG TC	
Salm-TEAQKS(R)	17	54.2	CGC TCT TTT GCG CCT CC	
StS-GQAIG(R)	17	57	CCG CCG ATT GCC TGA CC	
P3-16S	21	60.8	GGA ATA TTG CAC AAT GGG CGC	
1100R-16S	15	48	GGG TTG CGC TCG TTG	

DNA sequences are obtained for GUS genes from six different genera: *Enterobacter/Salmonella*, *Pseudomonas*, *Salmonella*, *Staphylococcus*, and *Thermotoga* (*see*, TIGR database at www.tigr.org) (Figures 4A-J and 16). Predicted amino acids translations are presented in Figures 3A-B and 17. In addition to the biochemical analysis and amplification using GUS primers, confirmation that the isolates contain a GUS gene is obtained from DNA and amino acid sequences. Amino acid alignment of *Bacillus* GUS (BGUS) with human (HGUS) and *E. coli* (EGUS) reveal extensive sequence identity and similarity. Likewise, alignment using ClustalW program of *Staphylococcus*, *Staphylococcus homini*, *Staphylococcus warneri*, *Thermotoga maritima*, *Enterobacter/Salmonella* and *E. coli*. show considerable amino acid identity and conservation (Figure 5B). The darker the shading, the higher the conservation among all GUSes. As seen in Figures 5B and 18, the region containing the critical catalytic residue (E344 using *Staphylococcus*_numbering) is highly conserved. This region extends over amino acids ca. 250 – ca. 360 and ca. 400 – ca. 535. Within these regions there are pockets of nearly complete identity. When constructing variants, in general, the regions of highest identity are not altered.

Two additional sequences from *Salmonella* and *Pseudomonas* are presented in nucleotide alignment with *Staphylococcus*. Significant sequence identity among the three sequences indicates that the *Salmonella* and *Pseudomonas* sequences are β-glucuronidase coding sequences. A full length *Salmonella* (CBP1) is also aligned with *E. coli* and *Staphylococcus* GUS. Overall identity is 71% and 51% nucleotide

identity to *E. coli* and *Staphylococcus*, respectively, and 85% and 46% amino acid identity to *E. coli* and *Staphylococcus*, respectively.

EXAMPLE 4

ISOLATION OF A GENE FROM *STAPHYLOCOCCUS* AND *SALMONELLA* ENCODING A SECRETED β-GLUCURONIDASE

Soil samples and skin samples are placed in broth and plated for growth of bacterial colonies on agar plates containing 50 µg/mL X-GlcA. Bacteria that secrete β-glucuronidase have a strong, diffuse staining pattern surrounding the colony.

One bacterial colony that exhibited this type of staining pattern is chosen. The bacterium is identified as a *Staphylococcus* based on amplification of 16S rRNA, and is most likely in the *Staphylococcus pseudomegaterium* group. Oligonucleotide sequences derived from areas exhibiting a high degree of similarity between *E. coli* and human β-glucuronidases are used in amplification reactions on *Staphylococcus* and *E. coli* DNA. A fragment is observed using *Staphylococcus* DNA, which is the same size as the *E. coli* fragment.

Staphylococcus DNA is digested with *Hind* III and ligated to *Hind* III-digested pBSII-KS plasmid vector. The recombinant plasmid is transfected into KW1, an *E. coli* strain that is deleted for the GUS operon. Cells are plated on X-GlcA plates, and one colony exhibited strong, diffuse staining pattern, suggesting that this clone encoded a secreted β-glucuronidase enzyme. The plasmid, pRAJa17.1, is isolated and subjected to analysis.

The DNA sequence of part of the insert of pRAJa17.1 is shown in Figure 1. A schematic of the 6029 bp fragment is shown in Figure 2. The fragment contains four large open reading frames. The open reading frame proposed as *Staphylococcus* GUS (GUS^{Stp}) begins at nucleotide 162 and extends to 1907 (Figure 1). The predicted translate is shown in Figure 3A and its alignment with *E. coli* and human β-glucuronidase is presented in Figure 5A. GUS^{Stp} is 47.2% identical to *E. coli* GUS,

which is about the same identity as human GUS and *E. coli* GUS (49.1%). Thus, GUS from *Staphylococcus* is about as related to another bacterium as to human. One striking difference in sequence among the proteins is the number of cysteine residues. Whereas, both human and *E. coli* GUS have 4 and 9 cysteines, respectively, GUS^{Stp} has only one cysteine.

The secreted GUS protein is 602 amino acids long and does not appear to have a canonical leader peptide. A prototypic leader sequence has an amino-terminal positively charged region, a central hydrophobic region, and a more polar carboxy-terminal region (*see*, von Heijne, *J. Membrane Biol. 115*:195-201, 1990) and is generally about 20 amino acids long. However, in both mammalian and bacterial cells, proteins without canonical or identifiable secretory sequences have been found in extracellular or periplasmic spaces.

A bacterium identified by 16SrRNA as Salmonella is isolated on the basis of halo formation. The predicted protein is 602 amino acids. There are 7 cysteine residues and 1 glycosylation site (Asn-Leu-Ser) at residue 358 (referenced to *E. coli* GUS). The *Salmonella* and *E. coli* sequences are very similar (71% nucleotide and 85% amino acid identity) reflecting the very close phylogeny of these genera. *Salmonella* GUS is less closely related to *Staphylococcus* GUS (51% nucleotide and 46% amino acid identity).

To simplify nomenclature, the following is proposed: the β-glucuronidase gene is called gusA. To distinguish origins of genes, a superscript is used to identify the genus, and species (if known). Thus *E. coli* GUS gene is $gusA^{Eco}$, *Staphylococcus* GUS gene is $gusA^{Stp}$, *Salmonella* GUS gene is $gusA^{Sal}$ and so on. Proteins are abbreviated as gus^{Eco}, GUS^{Stp} and so on.

EXAMPLE 5

PROPERTIES OF SECRETED β-GLUCURONIDASE

Although the screen described above suggests that the *Staphylococcus* GUS is secreted, the cellular localization of GUS^{Stp} is further examined. Cellular fractions (*e.g.*, periplasm, spheroplast, supernatant, etc.) are prepared from KW1 cells transformed with pRAJa17.1 or a subfragment that contains the GUS gene and from *E. coli* cells that express β-glucuronidase. GUS activity and β-galactosidase (β-gal) activity is determined for each fraction. The percent of total activity in the periplasm fraction for GUS and β-gal (a non-secreted protein) are calculated; the amount of β-gal activity is considered background and thus is subtracted from the amount of β-glucuronidase activity. In Figure 6, the relative activities of GUS^{Stp} and *E. coli* GUS in the periplasm fraction are plotted. As shown, approximately 50% of GUS^{Stp} activity is found in the periplasm, whereas less than 10% of *E. coli* GUS activity is present.

The thermal stability of GUS^{Stp} and *E. coli* GUS enzymes are determined at 65°C, using a substrate that can be measured by spectrophotometry, for example. One such substrate is p-nitrophenyl β-D-glucuronide (pNPG), which when cleaved by GUS releases the chromophore p-nitrophenol. At a pH greater than its pKa (approximately 7.15), the ionized chromophore absorbs light at 400-420 nm, therefore appears in the yellow range of visible light. Briefly, reactions are performed in 50 mM Na_3PO_4 pH 7.0, 10 mM 2-ME, 1 mM EDTA, 1 mM pNPG, and 0.1% Triton® X-100 at 37°C. The reactions are terminated by the addition of 0.4 ml of 2-amino-2-methylpropanediol, and absorbance measured at 415 nm against a substrate blank. Under these conditions, the molar extinction coefficient of p-nitrophenol is assumed to be 14,000. One unit is defined as the amount of enzyme that produces 1 nmole of product/min at 37°C.

As shown in Figure 7, GUS^{Stp} has a half-life of approximately 16 min, while *E. coli* GUS has a half-life of less than 2 min. Thus, GUS^{Stp} is at least 8 times more stable than the *E. coli* GUS. In addition, the catalytic properties of GUS^{Stp} are substantially better than the *E. coli* enzyme: The Km is approximately one-fourth to one-third and the Vmax is about the same at 37°C.

Table 2

	Staph GUS	*E. coli* GUS

Km	30-40 μM pNPG	120 μM pNPG
Vmax	80 nmoles/min/μg	80 nmoles/min/μg

The turnover number of GUS^{Stp} is approximately the same as *E.* coli GUS at 37°C and 2.5 to 5 times higher than *E. coli* GUS at room temperature (Figures 8 and 9). Turnover number is calculated as nmoles of pNPG converted to p-nitrophenol per min per μg of purified protein.

GUS^{Stp} enzyme activity is also resistant to inhibition by detergents. Enzyme activity assays are measured in the presence of varying amounts of SDS, Triton® X-100, or sarcosyl. As presented in Figure 10, GUS^{Stp} was not inhibited or only slightly inhibited (< 20% inhibition) in Triton® X-100 and Sarcosyl. In SDS, the enzyme still had substantial activity (60-75% activity). In addition, GUS^{Stp} is not inhibited by the end product of the reaction. Activity is determined normally or in the presence of 1 or 10 mM glucuronic acid. No inhibition is seen at either 1 or 10 mM glucuronic acid (Figure 11). The enzyme is also assayed in the presence of organic solvents, dimethylformamide (DMF) and dimethylsulfoxide (DMSO), and high concentrations of NaCl (Figure 12). Only at the highest concentrations of DMF and DMSO (20%) does GUS^{Stp} demonstrate inhibition, approximately 40% inhibited. In lesser concentrations of organic solvent and in the presence of 1 M NaCl, GUS^{Stp} retains essentially complete activity.

The *Staphylococcus* β-glucuronidase is secreted in *E. coli* when introduced in an expression plasmid as evidenced by approximately half of the enzyme activity being detected in the periplasm. In contrast, less than 10% of *E. coli* β-glucuronidase is found in periplasm. Secreted microbial GUS is also more stable than *E. coli* GUS (Figure 7), has a higher turnover number at both 37°C and room temperature (Figures 8 and 9), and unlike *E. coli* GUS, it is not substantially inhibited by detergents (Figure 10) or by glucuronic acid (Figure 11) and retains activity in high salt conditions and organic solvents (Figure 12).

As shown herein, multiple mutations at residues Val 128, Leu 141, Tyr 204 and Thr 560 (Figures 3A-B) result in a non-functional enzyme. Thus, at least

one of these amino acids is critical to maintaining enzyme activity. A mutein *Staphylococcus* GUS containing the amino acid alterations of Val 128 →Ala, Leu 141 →His, Tyr 204→Asp and Thr 560→Ala is constructed and exhibits little enzymatic activity. As shown herein, the residue alteration that most directly affected activity is Leu 141. In addition, three residues have been identified as likely contact residues important for catalysis in human GUS (residues Glu 451, Glu 540, and Tyr 504) (Jain *et al.*, *Nature Struct. Biol. 3*: 375, 1996). Based on alignment with *Staphylococcus* GUS, the corresponding residues are Glu 415, Glu 508, and Tyr 471. By analogy with human GUS, Asp 165 may also be close to the reaction center and likely forms a salt bridge with Arg 566. Thus, in embodiments where it is desirable to retain enzymatic activity of micorbial GUS, the residues corresponding to Leu 141, Glu 415, Glu 508, Tyr 471, Asp 165, and Arg 566 in *Staphylococcus* GUS are preferably unaltered.

EXAMPLE 6

CONSTRUCTION OF A CODON OPTIMIZED SECRETED β-GLUCURONIDASE

The *Staphylococcus* GUS gene is codon-optimized for expression in *E. coli* and in rice. Codon frequencies for each codon are determined by back translation using ecohigh codons for highly expressed genes of enteric bacteria. These ecohigh codon usages are available from GCG. The most frequently used codon for each amino acid is then chosen for synthesis. In addition, the polyadenylation signal, AATAAA, splice consensus sequences, ATTTA AGGT, and restriction sites that are found in polylinkers are eliminated. Other changes may be made to reduce potential secondary structure. To facilitate cloning in various vectors, four different 5' ends are synthesized: the first, called A0 (GT CGA CCC ATG GTA GAT CTG *ACT AGT* CTG TAC CCG) uses a sequence comprising an *Nco* I (underlined), *Bgl* II (double underlined), and *Spe* I (italicized) sites. The Leu (CTG) codon is at amino acid 2 in Figures 3A-B. The second variant, called AI (*GTC GAC* AGG AGT GCT ATC ATG CTG TAC CCG), adds the native Shine/Dalgarno sequence 5' of the initiator Met (ATG) codon; the third,

called AII, (*GTC GAC* AGG AGT GCT ACC ATG GTG TAC CCG) adds a modified Shine/Dalgarno sequence 5' of the initiator Met codon such that a *Nco* I site is added; the fourth one, called AIII (*GTC GAC* AGG AGT GCT ACC ATG GTA GAT CTG TAC CCG) adds a modified Shine/Dalgarno sequence 5' of the Leu (CTG) codon (residue 2) and *Nco* I and *Bgl* II sites.. All of these new 5' sequences contain a *Sal* I site at the extreme 5' end to facilitate construction and cloning. In certain embodiments, to facilitate protein purification, a sequence comprising a *Nhe* I, *Pml* I, and *Bst*E II sites (underlined) and encoding hexa-His amino acids joined at the 3' (COOH-terminus) of the gene.

GCTAGCCATCACCATCACCATCACGTGTGAATTGGTGACCG
SerSerHisHisHisHisHisHisVal *

Nucleotide and amino acid sequences of one engineered secretable microbial GUS are shown in Figures 13A-C, and a schematic is shown in Figure 14. The coding sequence for this protein is assembled in pieces. The sequence is dissected into four fragments, A (bases 1-457); B (bases 458-1012); C (bases 1013-1501); and D (bases 1502-1875). Oligonucleotides (Table 4) that are roughly 80 bases (range 36-100 bases) are synthesized to overlap and create each fragment. The fragments are each cloned separately and the DNA sequence verified. Then, the four fragments are excised and assembled in pLITMUS 39 (New England Biolabs, Beverley, MA), which is a small, high copy number cloning plasmid.

Table 3

Oligonucleotide	Size	Sequence	SEQ ID NO
gusA^{Stp} A-1-80T	80	TCGACCCATGGTAGATCTGACTAGTCTGTACCCGA TCAACACCGAGACCCGTGGCGTCTTCGACCTCAAT GGCGTCTGGA	
gusA^{Stp} A-121-200B	80	GGATTTCCTTGGTCACGCCAATGTCATTGTAACTG CTTGGGACGGCCATACTAATAGTGTCGGTCAGCTT GCTTTCGTAC	
gusA^{Stp} A-161-240T	80	CCAAGCAGTTACAATGACATTGGCGTGACCAAGGA AATCCGCAACCATATCGGATATGTCTGGTACGAAC GTGAGTTCAC	
gusA^{Stp} A-201-280B	80	GCGGAGCACGATACGCTGATCCTTCAGATAGGCCG GCACCGTGAACTCACGTTCGTACCAGACATATCCG ATATGGTTGC	

Oligonucleotide	Size	Sequence	SEQ ID NO
$gusA^{Stp}$ A-241-320T	80	GGTGCCGGCCTATCTGAAGGATCAGCGTATCGTGC TCCGCTTCGGCTCTGCAACTCACAAAGCAATTGTC TATGTCAATG	
$gusA^{Stp}$ A-281-360B	80	AATGGCAGGAATCCGCCCTTGTGCTCCACGACCAG CTCACCATTGACATAGACAATTGCTTTGTGAGTTG CAGAGCCGAA	
$gusA^{Stp}$ A-321-400T	80	GTGAGCTGGTCGTGGAGCACAAGGGCGGATTCCTG CCATTCGAAGCGGAAATCAACAACTCGCTGCGTGA TGGCATGAAT	
$gusA^{Stp}$ A-361-460B	100	GTACAGCCCCACCGGTAGGGTGCTATCGTCGAGGA TGTTGTCCACGGCGACGGTGACGCGATTCATGCCA TCACGCAGCGAGTTGTTGATTTCCGCTTCG	
$gusA^{Stp}$ A-401-456T	56	CGCGTCACCGTCGCCGTGGACAACATCCTCGACGA TAGCACCCTACCGGTGGGGCT	
$gusA^{Stp}$ A-41-120B	80	CACTTCTCTTCCAGTCCTTTCCCGTAGTCCAGCTT GAAGTTCCAGACGCCATTGAGGTCGAAGACGCCAC GGGTCTCGGT	
$gusA^{Stp}$ A-6-40B	35	TTGATCGGGTACAGACTAGTCAGATCTACCATGGG	
$gusA^{Stp}$ A-81-160T	80	ACTTCAAGCTGGACTACGGGAAAGGACTGGAAGAG AAGTGGTACGAAAGCAAGCTGACCGACACTATTAG TATGGCCGTC	
$gusA^{Stp}$ B-1-80T	80	GTACAGCGAGCGCCACGAAGAGGGCCTCGGAAAAG TCATTCGTAACAAGCCGAACTTCGACTTCTTCAAC TATGCAGGCC	
$gusA^{Stp}$ B-121-200B	80	CTTTGCCTTGAAAGTCCACCGTATAGGTCACAGTC CCGGTTGGGCCATTGAAGTCGGTCACAACCGAGAT GTCCTCGACG	
$gusA^{Stp}$ B-161-240T	80	ACCGGGACTGTGACCTATACGGTGGACTTTCAAGG CAAAGCCGAGACCGTGAAAGTGTCGGTCGTGGATG AGGAAGGCAA	
$gusA^{Stp}$ B-201-280B	80	CTCCACGTTACCGCTCAGGCCCTCGGTGCTTGCGA CCACTTTGCCTTCCTCATCCACGACCGACACTTTC ACGGTCTCGG	
$gusA^{Stp}$ B-241-320T	80	AGTGGTCGCAAGCACCGAGGGCCTGAGCGGTAACG TGGAGATTCCGAATGTCATCCTCTGGGAACCACTG AACACGTATC	
$gusA^{Stp}$ B-281-360B	80	GTCAGTCCGTCGTTCACCAGTTCCACTTTGATCTG GTAGAGATACGTGTTCAGTGGTTCCCAGAGGATGA CATTCGGAAT	
$gusA^{Stp}$ B-321-400T	80	TCTACCAGATCAAAGTGGAACTGGTGAACGACGGA CTGACCATCGATGTCTATGAAGAGCCGTTCGGCGT GCGGACCGTG	
$gusA^{Stp}$ B-361-440B	80	ACGGTTTGTTGTTGATGAGGAACTTGCCGTCGTTG ACTTCCACGGTCCGCACGCCGAACGGCTCTTCATA GACATCGATG	

Oligonucleotide	Size	Sequence	SEQ ID NO
gusA^{Stp} B-401-480T	80	GAAGTCAACGACGGCAAGTTCCTCATCAACAACAA ACCGTTCTACTTCAAGGGCTTTGGCAAACATGAGG ACACTCCTAT	
gusA^{Stp} B-41-120B	80	TACGTAAACGGGGTCGTGTAGATTTTCACCGGACG GTGCAGGCCTGCATAGTTGAAGAAGTCGAAGTTCG GCTTGTTACG	
gusA^{Stp} B-441-520B	80	ATCCATCACATTGCTCGCTTCGTTAAAGCCACGGC CGTTGATAGGAGTGTCCTCATGTTTGCCAAAGCCC TTGAAGTAGA	
gusA^{Stp} B-481-555T	75	CAACGGCCGTGGCTTTAACGAAGCGAGCAATGTGA TGGATTTCAATATCCTCAAATGGATCGGCGCCAAC AGCTT	
gusA^{Stp} B-5-40B	36	AATGACTTTTCCGAGGCCCTCTTCGTGGCGCTCGC T	
gusA^{Stp} B-521-559B	39	CCGGAAGCTGTTGGCGCCGATCCATTTGAGGATAT TGAA	
gusA^{Stp} B-81-160T	80	TGCACCGTCCGGTGAAAATCTACACGACCCCGTTT ACGTACGTCGAGGACATCTCGGTTGTGACCGACTT CAATGGCCCA	
gusA^{Stp} C-1-80T	80	CCGGACCGCACACTATCCGTACTCTGAAGAGTTGA TGCGTCTTGCGGATCGCGAGGGTCTGGTCGTGATC GACGAGACTC	
gusA^{Stp} C-121-200B	80	GTTCACGGAGAACGTCTTGATGGTGCTCAAACGTC CGAATCTTCTCCCAGGTACTGACGCGCTCGCTGCC TTCGCCGAGT	
gusA^{Stp} C-161-240T	80	ATTCGGACGTTTGAGCACCATCAAGACGTTCTCCG TGAACTGGTGTCTCGTGACAAGAACCATCCAAGCG TCGTGATGTG	
gusA^{Stp} C-201-280B	80	CGCGCCCTCTTCCTCAGTCGCCGCCTCGTTGGCGA TGCTCCACATCACGACGCTTGGATGGTTCTTGTCA CGAGACACCA	
gusA^{Stp} C-241-320T	80	GAGCATCGCCAACGAGGCGGCGACTGAGGAAGAGG GCGCGTACGAGTACTTCAAGCCGTTGGTGGAGCTG ACCAAGGAAC	
gusA^{Stp} C-281-360B	80	ACAAACAGCACGATCGTGACCGGACGCTTCTGTGG GTCGAGTTCCTTGGTCAGCTCCACCAACGGCTTGA AGTACTCGTA	
gusA^{Stp} C-321-400T	80	TCGACCCACAGAAGCGTCCGGTCACGATCGTGCTG TTTGTGATGGCTACCCCGGAGACGGACAAAGTCGC CGAACTGATT	
gusA^{Stp} C-361-440B	80	CGAAGTACCATCCGTTATAGCGATTGAGCGCGATG ACGTCAATCAGTTCGGCGACTTTGTCCGTCTCCGG GGTAGCCATC	
gusA^{Stp} C-401-489T	89	GACGTCATCGCGCTCAATCGCTATAACGGATGGTA CTTCGATGGCGGTGATCTCGAAGCGGCCAAAGTCC ATCTCCGCCAGGAATTTCA	

Oligonucleotide	Size	Sequence	SEQ ID NO
gusA^{Sp} C-41-120B	80	CCCGTGGTGGCCATGAAGTTGAGGTGCACGCCAAC TGCCGGAGTCTCGTCGATCACGACCAGACCCTCGC GATCCGCAAG	
gusA^{Sp} C-441-493B	53	CGCGTGAAATTCCTGGCGGAGATGGACTTTGGCCG CTTCGAGATCACCGCCAT	
gusA^{Sp} C-5-40B	36	ACGCATCAACTCTTCAGAGTACGGATAGTGTGCGG T	
gusA^{Sp} C-81-160T	80	CGGCAGTTGGCGTGCACCTCAACTTCATGGCCACC ACGGGACTCGGCGAAGGCAGCGAGCGCGTCAGTAC CTGGGAGAAG	
gusA^{Sp} D-1-80T	80	CGCGTGGAACAAGCGTTGCCCAGGAAAGCCGATCA TGATCACTGAGTACGGCGCAGACACCGTTGCGGGC TTTCACGACA	
gusA^{Sp} D-121-200B	80	TCGCGAAGTCCGCGAAGTTCCACGCTTGCTCACCC ACGAAGTTCTCAAACTCATCGAACACGACGTGGTT CGCCTGGTAG	
gusA^{Sp} D-161-240T	80	TTCGTGGGTGAGCAAGCGTGGAACTTCGCGGACTT CGCGACCTCTCAGGGCGTGATGCGCGTCCAAGGAA ACAAGAAGGG	
gusA^{Sp} D-201-280B	80	GTGCGCGGCGAGCTTCGGCTTGCGGTCACGAGTGA ACACGCCCTTCTTGTTTCCTTGGACGCGCATCACG CCCTGAGAGG	
gusA^{Sp} D-241-320T	80	CGTGTTCACTCGTGACCGCAAGCCGAAGCTCGCCG CGCACGTCTTTCGCGAGCGCTGGACCAACATTCCA GATTTCGGCT	
gusA^{Sp} D-281-369B	89	CGGTCACCAATTCACACGTGATGGTGATGGTGATG GCTAGCGTTCTTGTAGCCGAAATCTGGAATGTTGG TCCAGCGCTCGCGAAAGAC	
gusA^{Sp} D-321-373T	53	ACAAGAACGCTAGCCATCACCATCACCATCACGTG TGAATTGGTGACCGGGCC	
gusA^{Sp} D-41-120B	80	TACTCGACTTGATATTCCTCGGTGAACATCACTGG ATCAATGTCGTGAAAGCCCGCAACGGTGTCTGCGC CGTACTCAGT	
gusA^{Sp} D-5-40B	36	GATCATGATCGGCTTTCCTGGGCAACGCTTGTTCC A	
gusA^{Sp} D-81-160T	80	TTGATCCAGTGATGTTCACCGAGGAATATCAAGTC GAGTACTACCAGGCGAACCACGTCGTGTTCGATGA GTTTGAGAAC	

The AI form of microbial GUS in pLITMUS 39 is transfected into KW1 host *E. coli* cells. Bacterial cells are collected by centrifugation, washed with Mg salt solution and resuspended in IMAC buffer (50 mM Na_3PO_4, pH 7.0, 300 mM KCl, 0.1% Triton® X-100, 1 mM PMSF). For hexa-His fusion proteins, the lysate is clarified by centrifugation at 20,000 rpm for 30 min and batch absorbed on a Ni-IDA-Sepharose

column. The matrix is poured into a column and washed with IMAC buffer containing 75 mM imidazole. The β-glucuronidase protein bound to the matrix is eluted with IMAC buffer containing 10 mM EDTA.

If GUS is cloned without the hexa-His tail, the lysate is centrifuged at 50,000 rpm for 45 min, and diluted with 20 mM $NaPO_4$, 1 mM EDTA, pH 7.0 (buffer A). The diluted supernatant is then loaded onto a SP-Sepharose or equivalent column, and a linear gradient of 0 to 30% SP Buffer B (1 M NaCl, 20 mM $NaPO_4$, 1 mM EDTA, pH 7.0) in Buffer A with a total of 6 column volumes is applied. Fractions containing GUS are combined. Further purifications can be performed.

EXAMPLE 7

MUTEINS OF CODON OPTIMIZED β-GLUCURONIDASE

Muteins of the codon-optimized GUS genes are constructed. Each of the four GUS genes described above, A0, AI, AII, and AIII, contain none, one, or four amino acid alterations. The muteins that contain one alteration have a Leu 141 to His codon change. The muteins that contain four alterations have the Leu 141 to His change as well as Val 138 to Ala, Tyr 204 to Asp, and Thr 560 to Ala changes. pLITMUS 39 containing these 12 muteins are transfected into KW1. Colonies are tested for secretion of the introduced GUS gene by staining with X-GlcA. A white colony indicates undetectable GUS activity, a light blue colony indicates some detectable activity, and a dark blue colony indicates a higher level of detectable activity. As shown in Table 5 below, when GUS has the four mutations, no GUS activity is detectable. When GUS has a single Leu 141 to His mutation, three of the four constructs exhibit no GUS activity, while the AI construct exhibits a low level of GUS activity. All constructs exhibit GUS activity when no mutations are present. Thus, the Leu 141 to His mutation dramatically affects the activity of GUS.

Table 4

Number of Mutations	GUS construct			
	A0	AI	AII	AIII
4	white	white	white	white
1	white	light blue	white	white
0	light blue	dark blue	light blue	light blue

EXAMPLE 8

EXPRESSION OF MICROBIAL β-GLUCURONIDASES IN YEAST, PLANTS AND *E. COLI*

A series of expression vector constructs of three different GUS genes, *E. coli* GUS, *Staphylococcus* GUS, and the A0 version of codon-optimized *Staphylococcus* GUS, are prepared and tested for enzymatic activity in *E. coli*, yeast, and plants (rice, Millin variety). The GUS genes are cloned in vectors that either contain a signal peptide suitable for the host or do not contain a signal peptide. The *E. coli* vector contains a sequence encoding a *pel*B signal peptide, the yeast vectors contain a sequence encoding either an invertase or Mat alpha signal peptide, and the plant vectors contain a sequence encoding either a glycine-rich protein (GRP) or extensin signal peptide.

Invertase signal sequence:
ATGCTTTTGC AAGCCTTCCT TTTCCTTTTG GCTGGTTTTG CAGCCAAAAT ATCTGCAATG (SEQ ID NO._____)

Mat alpha signal sequence:
ATGAGATTTC CTTCAATTTT TACTGCAGTT TTATTCGCAG CATCCTCCGC ATTAGCTGCT
CCAGTCAACA CTACAACAGA AGATGAAACG GCACAAATTC CGGCTGAAGC TGTCATCGGT
TACTTAGATT TAGAAGGGGA TTTCGATGTT GCTGTTTTGC CATTTTCCAA CAGCACAAAT
AACGGGTTAT TGTTTATAAA TACTACTATT GCCAGCATTG CTGCTAAAGA AGAAGGGGTA
TCTTTGGATA AAAGAGAG (SEQ ID NO.____)

Extensin signal sequence
CATGGGAAAA ATGGCTTCTC TATTTGCCAC ATTTTTAGTG GTTTTAGTGT CACTTAGCTT
AGCTTCTGAA AGCTCAGCAA ATTATCAA (SEQ ID NO.____)

GRP signal sequence
CATGGCTACT ACTAAGCATT TGGCTCTTGC CATCCTTGTC CTCCTTAGCA TTGGTATGAC
CACCAGTGCA AGAACCCTCC TA (SEQ ID NO._____)

The GUS genes are cloned into each of these vectors using standard recombinant techniques of isolation of a GUS-gene containing fragment and ligation into an appropriately restricted vector. The recombinant vectors are then transfected into the appropriate host and transfectants are tested for GUS activity.

As shown in the Table below, all tested transfectants exhibit GUS activity (indicated by a +). Moreover, similar results are obtained regardless of the presence or absence of a signal peptide.

Table 5

GUS	*E. coli*		Yeast			Plants		
	No SP*	pelB	No SP	Invertase	Mat α	No SP	GRP	Extensin
E. coli GUS	+	NT	+	+	+	+	+	+
Staphylococcus GUS	+	NT	+	+	+	+	+	+

*; SP=signal peptide

EXAMPLE 9

ELIMINATION OF THE POTENTIAL N-GLYCOSYLATION SITE OF *STAPHYLOCOCCUS* β-GLUCURONIDASE

The consensus N-glycosylation sequence Asn-X-Ser/Thr is present in *Staphylococcus* GUS at amino acids 118-120, Asn-Asn-Ser (Figures 3A-B). Glycosylation could interfere with secretion or activity of β-glucuronidase upon entering the ER. To remove potential N-glycosylation, the Asn at residue 118 is changed to another amino acid in the plasmid pTANE95m (AI) is altered. The GUS in this plasmid is a synthetic GUS gene with a completely native 5' end.

The oligonucleotides Asn-T, 5'-A TTC CTG CCA TTC GAG GCG GAA ATC NNG AAC TCG CTG CGT GAT-3' (SEQ ID No. ___) and Asn-B, 5'-ATC ACG CAG CGA GTT CNN GAT TTC CGC CTC GAA TGG CAG GAA T-3' (SEQ ID No. ____), are used in the "quikchange" mutagenesis method by Stratagene (La

Jolla, CA) to randomize the first two nucleotides of the Asn 118 codon, AAC. The third base is changed to a G nucleotide, so that reversion to Asn is not possible. In theory a total of 13 different amino acids are created at position 118.

Because expression of GUS from the plasmid pTANE95m (AI) exhibits a range of colony phenotypes from white to dark blue, a restriction enzyme digestion assay is used to confirm presence of mutants. Therefore, an elimination of a *Bst*B I restriction site which does not change any amino acid, is also introduced into the mutagenizing oligonucleotides to facilitate restriction digestion screening of mutants.

Sixty colonies were randomly picked and assayed by *Bst*B I digestion. Twenty-one out of the 60 colonies have the *Bst*B I site removed and are thus mutants. DNA sequence analysis of these candidate mutants show that a total of 8 different amino acids are obtained. Five of the N118 mutants are chosen as suitable for further experimentation. In these mutants, the N118 residue is changed to a Ser, Arg, Leu, Pro, or Met.

EXAMPLE 10

EXPRESSION OF β-GLUCURONIDASE IN TRANSGENIC RICE PLANTS

Microbial GUS can be used as a non-destructible marker. In this example, transgenic rice expressing a GUS gene encoding a secreted form are assayed for GUS expression *in planta*.

Seeds of T0 plants, which are the primary transformed plants, from pTANG86.1/2/3/4/5/6 (see Table 7 below) transformed plants, seeds of pCAM1301 (*E. coli* GUS with N358-Q change to remove N-glycosylation signal sequence) transformed plants, or untransformed Millin rice seeds are germinated in water containing 1 mM MUG or 50 μg/mL X-GlcA with or without hygromycin (for nontransformed plants). Resulting plants are observed for any reduced growth due to the presence of MUG, X-GlcA. No toxic effects of X-GlcA are detected, but roots of the plants grown in MUG are somewhat stunted.

For assaying GUS activity *in planta*, seeds are germinated in water with or without hygromycin (for nontransformed plants). Roots of the seedlings are submerged in water containing 1 mM MUG, or 50 µg/mL X-GlcA. Fluorescence (in the case of MUG staining) or indigo dye (in the case of X-GlcA staining) are assayed in the media and roots over time.

Secondary roots from seedlings of pTANG86.3 and pTANG86.5 (GUS^{Stp} fused with signal peptides) plants show indigo color after ½ hour incubation in water containing X-GlcA. Evidence that GUS is a non-destructive marker is obtained by plant growth after transferring the stained plant to water. Furthermore, stained roots also grow further.

EXAMPLE 11

EXPRESSION OF β-GLUCURONIDASE IN YEAST

All the yeast plasmids are based on the Ycp backbone, which contains a yeast centromere and is stable at low copy number. Yeast strain InvSc1 (*mat* α *his*3-Δ1 *leu*2 *trp*1-289 *ura*3-52) from Invitrogen (Carlsbad, CA) is transformed with the *E. coli* GUS and *Staphylococcus* GUS plasmids indicated in the table below. Transformants are plated on both selection media (minimal media supplemented with His, Leu, Trp, and 2% glucose as a carbon source to suppress the expression of the gene driven by the *gal*1 promoter) and expression media (media supplemented with His, Leu, Trp, 1% raffinose, 1% galactose as carbon source and with 50 µg/ml X-GlcA).

Table 6

	Yeast			Plants		
	No SP	Invertase	Mat alpha	No SP	GRP	Extensin
E. coli	pAKD80.3	pAKD80.6	pTANG87.4	pTANG86.2	pTANG86.4	pTANG86.6
Syn BGUS	pTANG87.1	pTANG87.2	pTANG87.3	pTANG86.1	pTANG86.3	pTANG86.5
Nat BGUS	pAKD102.1	pAKE2.1	pAKE11.4	pAKD40	pAKC30.1	pAKC30.3

With the exception of pAKD80.6, all other transformed yeast colonies are white on X-GlcA plates. The transformants do express GUS, however, which is evidenced by lysing the cells on the plates with hot agarose containing X-GlcA and observing the characteristic indigo color. The yeast transformants are white when GUS is not secreted, as X-GlcA cannot be taken by the yeast cell. All the yeast colonies transformed with pAKD80.6 are blue on X-GlcA plates and have a blue halo around each colony, clearly indicating that the enzyme is secreted into the medium.

Staphylococcus GUS enzyme has a potential N-glycosylation site, which may interfere with the secretion process or cause inactivation of the enzyme upon secretion. To determine whether the N-glycosylation site has a deleterious effect, on secretion, yeast colonies are streaked on expression plates containing X-GlcA and from 0.1 to 20 μg/ml of tunicamycin (to inhibit all N-glycosylation). At high concentrations of tunicamycin (5, 10, and 20 μg/ml), yeast colonies do not grow, likely due to toxicity of the drug. However, in yeast transformed with pTANG87.3, the cells that do survive at these tunicamycin concentrations are blue. This indicates that glycosylation may affect the secretion or activity of *Staphylococcus* GUS. Any effect should be overcome by mutating the glycosylation signal sequence as described.

EXAMPLE 12

EXPRESSION OF LOW-CYSTEINE *E. COLI* β-GLUCURONIDASE

The *E. coli* GUS protein has nine cysteine residues, whereas, human GUS has four and *Staphylococcus* GUS has one. Low-cysteine muteins of *E. coli* GUS are constructed to provide a form of *Ec*GUS that is secretable.

Single and multiple Cys muteins are constructed by site-directed mutagenesis techniques. Eight of the nine cysteine residues in *E. coli* GUS are changed to the corresponding residue found in human GUS based on alignment of the two protein sequences. One of the *E. coli* GUS cysteine residues, amino acid 463, aligns with a cysteine residue in human GUS and was not altered. The corresponding amino acids between *E. coli* GUS and human GUS are shown below.

Table 7

Identifier	EcGUS Cys residue no.	Human GUS corresponding amino acid
A	28	Asn
B	133	Ala
C	197	Ser
D	253	Glu
E	262	Ser
F	442	Phe
G	448	Tyr
H	463	Cys
I	527	Lys

The mutein GUS genes are cloned into a pBS backbone. The mutations are confirmed by diagnostic restriction site changes and by DNA sequence analysis. Recombinant vectors are transfected into KW1 and GUS activity assayed by staining with X-GlcA (5-bromo-4-chloro-3-indolyl-β-D-glucuronide).

As shown in the Table below, when the Cys residues at 442 (F), 448 (G), and 527 (I) are altered, GUS activity is greatly or completely diminished. In contrast,

when the N-terminal five Cys residues (A, B, C, D, and E) are altered, GUS activity remains detectable.

Table 8

Cys changes	GUS activity
A	Yes
B	Yes
C	Yes
I	No
D, E	Yes
F, G	No
C, D, E	Yes
B, C, D, E	Yes
A, B, C, D, E	Yes
A, B, C, D, E, I	No

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.